

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

RAWLS ET L'UTILITARISME ;
UNE ANALYSE DE L'UTILISATION DE LA THÉORIE DU CHOIX RATIONNEL
DANS LES THÉORIES DE JOHN RAWLS ET DE JOHN HARSANYI

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN PHILOSOPHIE

PAR
MIKAEL ROY-THOUIN

DÉCEMBRE 2008

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Tout d'abord, je dois remercier Mme Jocelyne Couture, professeure à l'UQAM et directrice de ce projet de mémoire. Ses conseils et ses critiques, toujours aussi clairs et pertinents les uns que les autres, ont su orienter efficacement mes lectures et guider judicieusement l'argumentation présentée dans mon texte. Je remercie également Mme Christine Strachle et M. Serge Robert pour leurs critiques et commentaires fort instructifs.

Je veux également remercier Mme Idil Boran, aujourd'hui professeure à l'Université York, qui m'a enseigné à deux reprises lors de son passage à l'UQAM. Tout d'abord dans le cadre d'un cours de premier cycle portant sur l'éthique économique et sociale, puis dans celui d'un séminaire de deuxième cycle portant sur le contractualisme. La rencontre avec Mme Boran a été particulièrement importante pour moi puisqu'elle m'a permis de constater que les questions qui suscitent le plus mon intérêt sont bel et bien discutées de nos jours dans les départements de philosophie.

Par ailleurs, je tiens à remercier M. Clément Lemelin, ex-professeur d'économie à l'École des sciences de la gestion de l'UQAM aujourd'hui à la retraite. Il y a maintenant près de dix ans, M. Lemelin a été le premier à me parler de John Rawls et de ses deux principes de justice. Il a aussi été le premier à me mentionner le nom de John Harsanyi. À l'époque, j'étais loin de me douter que j'allais étudier ces deux auteurs dans le cadre d'un mémoire de maîtrise en philosophie. Aujourd'hui, il me serait impossible de nier l'influence déterminante qu'a exercée M. Lemelin sur mon parcours universitaire.

Merci également à Josie, pour son aide précieuse relativement aux dernières corrections de ce mémoire. Finalement, un merci tout spécial à mes parents, Danièle et André. Depuis ma tendre enfance, tous deux ont su encourager ma curiosité et susciter mon goût du savoir. De plus, les valeurs qu'ils m'ont transmises ne sont certes pas étrangères aux thèmes que j'aborde dans ce mémoire et à la façon que j'ai de les traiter.

TABLE DES MATIÈRES

RÉSUMÉ.....	v
INTRODUCTION.....	1
CHAPITRE I: LA THÉORIE DE LA JUSTICE COMME ÉQUITÉ DE JOHN RAWLS	6
1.1 Introduction	6
1.2 Philosophie morale et justification : la méthode de l'équilibre réfléchi.....	7
1.3 Justification publique, pluralisme raisonnable et consensus par recoupement	10
1.4 Le contractualisme de Rawls.....	12
1.5 Les circonstances de la justice.....	15
1.6 Les deux principes de justice.....	16
1.7 La position originelle	18
1.8 Le rationnel et le raisonnable	22
1.9 Les arguments en faveur de la règle du maximin.....	24
1.10 Conclusion.....	28
CHAPITRE II: LA THÉORIE MORALE UTILITARISTE DE JOHN HARSANYI.....	30
2.1 Introduction	30
2.2 L'évolution historique du concept d'utilité	32
2.3 L'utilitarisme expérientialiste et l'utilitarisme préférentialiste	39
2.4 Le modèle d'équiprobabilité pour les jugements moraux	41
2.5 L'utilitarisme de l'acte et l'utilitarisme de la règle	45
2.6 Le désaccord entre Rawls et Harsanyi	48
2.7 Conclusion.....	51
CHAPITRE III: LA RELATION ENTRE LA THÉORIE DE RAWLS ET L'UTILITARISME.....	53
3.1 Introduction	53
3.2 Rawls, un utilitariste?.....	54
3.3 Le semi-conséquentialisme de Rawls.....	59

3.4 La possibilité d'une justification utilitariste des principes de justice de Rawls	61
3.5 Les deux niveaux de justification dans la théorie de la justice de Rawls	65
3.6 Le contractualisme de Rawls et la position originelle	70
3.7 Motivations internes et raisons externes	75
3.8 Affinités et points de désaccords entre la théorie de Rawls et l'utilitarisme	77
3.9 Différentes conceptions de la rationalité	80
3.10 Conclusion	82
CONCLUSION	84
RÉFÉRENCES	90
BIBLIOGRAPHIE SÉLECTIVE	93

RÉSUMÉ

Depuis la parution de *A Theory of Justice* en 1971, la théorie de la justice comme équité de l'auteur américain John Rawls occupe une place importante dans le domaine de la justice économique et sociale. Les deux principes de justice qu'elle énonce, celui de la plus grande liberté égale pour tous ainsi que le principe de différence associé à une juste égalité des chances, ont été et continuent d'être largement discutés, analysés et critiqués aujourd'hui. Ces principes, applicables dans le cadre d'une démocratie constitutionnelle contemporaine caractérisé par le fait du pluralisme raisonnable, sont présentés par Rawls comme étant ceux que choisirait rationnellement une personne placée derrière un voile d'ignorance l'empêchant de savoir quelle position elle occupe dans la société. Cette situation initiale hypothétique dans laquelle s'effectue le choix des principes de justice par un individu représentatif constitue ce que Rawls décrit comme étant la position originelle.

Dans des circonstances similaires à celles de cette position originelle, John Harsanyi estime quant à lui que le choix d'un individu rationnel n'aboutirait pas aux deux principes proposés par Rawls mais plutôt au principe de la maximisation de l'utilité. Pour en arriver à cette conclusion, Harsanyi utilise la théorie bayésienne de la décision rationnelle. Or, alors que la théorie bayésienne implique un recours abondant à des raisonnements probabilistes, Rawls rejette ce type de raisonnement probabiliste au sein de la position originelle. Il appert ainsi que Rawls et Harsanyi utilisent différemment la théorie du choix rationnel en situation d'incertitude, ce qui les amène à défendre des théories de la justice différentes : d'un côté, Rawls élabore une théorie déontologique; de l'autre, Harsanyi développe une version particulièrement sophistiquée de la théorie conséquentialiste qu'est l'utilitarisme. À mes yeux, l'utilisation faite par Harsanyi est en meilleure symbiose avec une application conventionnelle de la théorie de la décision rationnelle.

Par ailleurs, bien que Rawls présente ses deux principes comme le résultat du choix rationnel d'un seul individu représentatif, l'idée principale derrière sa théorie de la justice est celle d'un accord volontaire portant sur les termes d'une coopération sociale mutuellement avantageuse entre personnes libres, égales et rationnelles. Je soutiens que cette idée contractualiste est mal servie par l'argumentation à partir de la position originelle et que les véritables arguments de Rawls en faveur de ses deux principes de justice et contre le principe utilitariste sont plus efficaces lorsqu'ils sont présentés indépendamment de la construction de la position originelle. En outre, alors que Rawls présentait initialement sa théorie comme un concurrent direct de l'utilitarisme, la relation entre les deux théories mérite d'être nuancée et précisée à la lumière des écrits subséquents de l'auteur de *Théorie de la Justice*. Ainsi, lorsque l'on examine leurs applications dans un contexte réaliste, l'utilitarisme et la théorie de la justice comme équité apparaissent en bout de ligne davantage comme des alliés que comme des adversaires sur l'échiquier politique des démocraties libérales contemporaines.

Mots-clés : Rawls, Harsanyi, principes de justice, utilitarisme, théorie du choix rationnel.

INTRODUCTION

Comment détermine-t-on ce qui est « bien » et ce qui est « mal »? Qu'est-ce qu'une action « bonne »? Quel comportement doit adopter quelqu'un voulant « bien » agir? Voilà les principales questions générales auxquelles la philosophie morale, ou autrement dit l'éthique, essaie de fournir des réponses pertinentes depuis des millénaires. Cela dit, des questions plus pointues sont aujourd'hui posées par une branche relativement nouvelle de l'éthique, celle de la justice économique et sociale. Quelles sont les « bonnes » règles pour encadrer l'interaction sociale? Des individus formant une société ont-ils des droits et des devoirs fondamentaux? Si oui, quels sont-ils? Qu'est-ce qu'une répartition « juste » des ressources rares? La justice sociale commande-t-elle une redistribution de la richesse? Jusqu'à quel point? Toutes ces questions, et bien d'autres, sont au cœur de l'analyse philosophique contemporaine portant sur la justice des institutions sociales dans le cadre d'un régime politique démocratique. C'est sur ce thème que porte le présent mémoire, qui s'intéresse plus particulièrement à la conception politique de la justice de John Rawls, appelée théorie de la justice comme équité, ainsi qu'à la théorie utilitariste de John Harsanyi.

Malgré certaines distinctions importantes à faire entre leurs théories, Rawls et Harsanyi s'entendent pour dire que les principes justes pour régir le fonctionnement des institutions sociales sont ceux que choisirait une personne rationnelle ne sachant pas quelle position elle occupe dans la société. Tous deux font donc appel à la théorie du choix rationnel en situation d'incertitude, aussi appelée théorie de la décision rationnelle, pour expliquer et justifier le choix de principes de justice. Or, alors que l'analyse d'Harsanyi mène à la conclusion que c'est le principe d'utilité moyenne qui doit être choisi, Rawls argumente en faveur du choix de deux principes de justice garantissant d'une part un système de droits et libertés le plus étendu possible et offrant, d'autre part, un critère pour déterminer l'étendue des inégalités qui devraient être tolérées au nom de l'intérêt de chacun dans le cadre d'une juste égalité des chances. Ainsi, il appert que Rawls et Harsanyi ne s'entendent pas sur la

façon d'utiliser la théorie de la décision rationnelle dans le contexte du choix de principes de justice. En effet, tandis qu'Harsanyi s'appuie sur la théorie bayésienne de la décision rationnelle et fait un usage abondant de raisonnements probabilistes, Rawls rejette la théorie bayésienne et écarte tout raisonnement probabiliste de son analyse. L'un des principaux objectifs de ce mémoire est d'expliquer et de clarifier la nature du désaccord entre Rawls et Harsanyi sur ce point.

L'œuvre de Rawls étant colossale et fort complexe dans son ensemble, il est évident qu'il m'est ici impossible de traiter en détails de tous les arguments philosophiques qu'elle contient. Mon analyse des travaux de Rawls se concentrera sur l'utilisation qu'il fait de la théorie du choix rationnel dans *Théorie de la justice*, son premier ouvrage de grande envergure. C'est donc principalement à l'argumentation déployée par Rawls à partir de la fameuse position originelle en faveur de ses deux principes de justice et contre le principe utilitariste que je m'intéresserai. Cela dit, malgré l'emphasis mise sur cette partie de la théorie de Rawls, il ne faut jamais perdre de vue qu'elle n'en constitue qu'un des trois pans. Car en effet, si les principes de justice rawlsiens sont en bonne partie justifiés par le choix rationnel d'un individu ignorant la position sociale qu'il occupe, cette justification ne saurait être complète sans tenir compte des implications de l'application des principes et de la stabilité qu'ils peuvent apporter aux institutions sociales. Je serai ainsi forcé de faire les liens qui s'imposent entre ces différents éléments, mais c'est néanmoins sur le premier que je me concentrerai. Ce faisant, j'espère apporter un certain éclairage sur l'utilisation que fait Rawls de la théorie de la décision rationnelle tout en m'efforçant de rendre justice à l'œuvre complète de Rawls et en demeurant fidèle aux idées qui fondent sa théorie de la justice.

La littérature secondaire sur Rawls étant abondante et fort variée, je me dois encore une fois de faire des choix quant aux commentateurs auxquels je ferai référence. Dans le contexte d'une analyse de l'utilisation de la théorie du choix rationnel servant à justifier des principes de justice, les travaux de John Harsanyi me semblent particulièrement intéressants à comparer avec l'œuvre de Rawls pour deux principales raisons. Premièrement, comme je l'ai déjà mentionné, même s'il part de prémisses semblables, Harsanyi n'utilise pas la théorie de la décision rationnelle de la même manière que Rawls. Deuxièmement, Harsanyi tente de

justifier à l'aide de la théorie de la décision rationnelle le choix du principe d'utilité moyenne pour guider le fonctionnement des institutions sociales alors que c'est justement en tant que doctrine concurrente de la doctrine utilitariste que Rawls avait, au départ, présenté sa théorie de la justice comme équité. Cela dit, la théorie d'Harsanyi constitue une version sophistiquée et remplie de subtilités qui s'écarte en certains points importants de la stricte doctrine utilitariste classique, ce qui rend la comparaison avec la théorie de Rawls plus difficile, mais d'autant plus intéressante.

Je soutiendrai qu'il y a une affinité naturelle entre l'utilitarisme et la théorie du choix rationnel. Je souhaite montrer qu'un utilitarisme de la règle, tel que développé par John Harsanyi, est pleinement compatible avec le cadre traditionnellement fourni par la théorie du choix rationnel et, en ce sens bien précis, s'avère supérieur à la théorie de Rawls. Bien entendu, il existe une multitude d'autres aspects en regard desquels la théorie de Rawls peut être jugée supérieure à celle d'Harsanyi. Je veux simplement avancer l'idée que l'utilitarisme de la règle est en meilleure symbiose avec une application conventionnelle et rigoureuse de la théorie du choix rationnel au contexte de la position originelle que ne l'est la théorie de la justice comme équité. Cela n'enlève rien par ailleurs à la force de l'ensemble des arguments philosophiques mis de l'avant par Rawls pour défendre ses deux principes de justice. Cependant, une relecture en des termes utilitaristes de l'argumentation à partir de la position originelle suggère que des arguments plus directs en faveur des deux principes pourraient s'avérer plus efficaces et pertinents que ne l'est finalement l'argumentation indirecte de Rawls, qui repose sur une construction artificielle du contexte de choix hypothétique dans lequel sont placés les partenaires ainsi que sur la sélection des contraintes qui doivent être imposées à ce choix pour qu'il aboutisse au résultat souhaité.

La comparaison entre la théorie de Rawls et celle d'Harsanyi fournira par ailleurs la base d'une analyse de la relation complexe qu'entretient la théorie de Rawls avec l'utilitarisme en général et, plus précisément, avec différentes formes de la doctrine utilitariste. Pour nuancer et préciser cette relation, je m'appuierai sur des écrits de Sydney Alexander, Thomas Pogge, T.M. Scanlon, Brian Barry, Catherine Audard et Samuel Scheffler. Plusieurs de ces auteurs ayant abondamment critiqué et commenté l'œuvre de

Rawls, j'ai ici aussi choisi de me concentrer principalement sur les éléments liés à l'utilisation de la théorie de la décision rationnelle. Toutefois, je ferai également certains liens inévitables avec des arguments liés à d'autres considérations, notamment celles ayant trait à la stabilité des institutions au sein d'une société pluraliste et démocratique. En bout de ligne, je soutiendrai que Rawls ne parvient pas à rejeter de manière définitive l'approche utilitariste en tant que doctrine servant de cadre à l'élaboration d'une théorie morale fondée sur la rationalité des individus. Et ce, même s'il parvient par de solides arguments à démontrer que le principe selon lequel la maximisation de l'utilité doit servir directement de guide au fonctionnement des institutions ne représente pas le meilleur choix que peuvent faire des agents rationnels soumis à certaines conditions idéales. D'où une possible réconciliation entre l'utilitarisme et la théorie de la justice comme équité¹, ce qui est par ailleurs souhaitable pour que les utilitaristes puissent faire partie du consensus par recoupement recherché par Rawls. Cela dit, je réfuterai la thèse réductrice selon laquelle la théorie de Rawls n'est ni plus ni moins qu'un utilitarisme indirect et montrerai ce qui distingue la théorie de la justice comme équité des théories utilitaristes.

Mon texte se divise en trois chapitres. Le premier servira à présenter la théorie de la justice comme équité de John Rawls. Une attention particulière y sera accordée à l'argumentation à partir de la position originelle, qui fait usage de la théorie du choix rationnel en situation d'incertitude. Le deuxième chapitre se penchera sur la théorie utilitariste de John Harsanyi, qui présente une manière originale de justifier le choix du principe d'utilité moyenne. Enfin, le troisième chapitre sera consacré à clarifier la relation entre la théorie de Rawls et l'utilitarisme. Aux termes de ces trois chapitres, j'espère que les arguments fondamentaux de Rawls pour critiquer l'utilitarisme ressortiront plus clairement. Je souhaite en outre montrer que ces arguments peuvent être avancés indépendamment de la

¹ La réconciliation peut aller bien au-delà de l'utilitarisme et de la justice comme équité puisque, dans une perspective beaucoup plus large, ce sont les deux grandes écoles de pensée rivales en philosophie morale, i.e. l'approche conséquentialiste et l'approche déontologique, qui peuvent être incorporées en une seule. Cela étant dit, de nombreux défenseurs de l'approche déontologique résisteront sans doute en affirmant, non sans tort, que cette réconciliation se fait au prix de l'obscurcissement de certaines questions philosophiques importantes sur la nature de la morale et de son emprise sur l'esprit humain. D'un autre côté, dans quelle mesure peut-on réellement qualifier de purs utilitaristes des gens qui mettent finalement de l'avant d'autres principes de justice que le principe utilitariste?

construction de la position originelle et qu'une argumentation contractualiste plus directe en faveur des principes de justice rawlsiens est plus fructueuse que l'argumentation proposée par Rawls à partir de la position originelle. Une des raisons principales qui appuie cette prise de position est issue du refus de Rawls de recourir à la théorie bayésienne de la décision rationnelle. J'estime en effet que la controverse entourant ce refus contribue à embrouiller les véritables arguments de Rawls contre l'utilitarisme. Enfin, je souhaite montrer que, malgré certaines différences fondamentales, l'utilitarisme et la théorie de la justice comme équité ne sont pas nécessairement en opposition stricte dans le contexte de l'élaboration de principes de justice régissant le fonctionnement des institutions sociales. Au contraire, il me semble que les deux théories doivent pouvoir se rencontrer afin que les motivations individuelles puissent coïncider avec des raisons morales d'agir, malgré les difficultés que cela comporte.

CHAPITRE I

LA THÉORIE DE LA JUSTICE COMME ÉQUITÉ DE JOHN RAWLS

1.1 Introduction

À une époque où les sciences humaines et sociales semblent se compartimenter de plus en plus et où l'approche analytique domine largement au sein de la communauté des philosophes chercheurs, l'approche de l'auteur américain John Rawls (1921-2002) s'inscrit en faux. Renouant avec la tradition des principaux penseurs modernes de la théorie du contrat social (Hobbes, Locke, Rousseau, Kant), l'œuvre de Rawls, de par son envergure et son caractère multidisciplinaire, peut également être vue comme appartenant à la lignée des grands écrits de l'économie politique classique (Smith, Ricardo, Mill, etc.). Relativement facile d'accès pour des non-spécialistes malgré toute sa richesse et sa complexité, *A Theory of Justice*, ouvrage monumental paru en 1971 en version originale anglaise et en 1987 en version française (*Théorie de la justice*), a eu un impact majeur immédiat tant au sein des cercles de chercheurs universitaires, des hommes politiques que parmi le grand public en général. Notons également la publication des livres *Political Liberalism* en 1993 (v.f. : *Libéralisme politique* en 1995) et *The Law of Peoples* en 1999, auxquels s'ajoutent des dizaines d'articles publiés tout au long de sa prolifique carrière dans de revues spécialisées de philosophie, de politique, de droit et d'économie. L'influence de Rawls est aujourd'hui telle qu'il est difficile de traiter de philosophie politique sans tenir compte de son œuvre et pratiquement impossible de parler des théories de la justice sociale sans à tout le moins mentionner son nom. En effet, peu importe que l'on souhaite aligner sa position sur la sienne ou défendre des thèses qui sont opposés ou contradictoires à sa théorie, ignorer Rawls est maintenant impensable lorsqu'il est question de justice distributive.

L'œuvre majeure de Rawls, *Théorie de la justice*, sera au cœur de mon analyse de la théorie de la justice comme équité dans les prochaines pages. Plus particulièrement, la troisième section du livre, qui porte sur la position originelle, sera analysée et discutée abondamment vers la fin du chapitre. Je retracerai alors l'argumentation qu'y présente Rawls en faveur des deux principes de justice et contre le principe utilitariste. Mais avant de me lancer dans le vif du sujet, j'emprunterai quelques détours nécessaires. En effet, l'examen des points que je présenterai préalablement s'avère indispensable pour rendre justice à toute la richesse et la subtilité de la théorie de Rawls et mieux situer l'analyse que j'en fais. Ainsi, je commencerai par clarifier certaines questions d'ordre méthodologique en exposant la vision qu'a Rawls du type de justification appropriée en philosophie morale en présentant succinctement la méthode de l'équilibre réfléchi. Ceci m'amènera ensuite à présenter trois concepts importants du libéralisme politique dont se réclame Rawls : la justification publique, le pluralisme raisonnable et le consensus par recoupement. Puis, je traiterai de manière générale du contractualisme de Rawls et de l'idée principale qui sous-tend sa conception politique de la justice, c'est-à-dire celle d'un accord volontaire entre personnes libres, égales et rationnelles. J'examinerai ensuite brièvement les circonstances objectives et subjectives de la justice décrites par Rawls, qui fut largement inspiré par David Hume sur ce point. S'ensuivra une présentation des deux principes de justice mis de l'avant par Rawls et de la règle de priorité lexicographique. Finalement, je traiterai en détails de l'argumentation à partir de la position originelle servant à défendre ces principes et à rejeter le principe utilitariste. En plus de décrire la position originelle ainsi que les partenaires qui s'y retrouvent, je retracerai le raisonnement des partenaires en insistant sur la distinction faite par Rawls entre le Rationnel et le Raisonnable. Les arguments en faveur de la règle de décision du maximin au détriment de celle de la maximisation de l'utilité recevront également une attention particulière.

1.2 Philosophie morale et justification : la méthode de l'équilibre réfléchi

D'inspiration résolument socratique, la méthode de l'équilibre réfléchi est celle que privilégie Rawls pour la pratique de la philosophie morale. La recherche d'un équilibre

réfléchi implique notamment que toute croyance, des principes les plus généraux aux jugements factuels les plus particuliers, est éventuellement susceptible d'être révisé. Cela étant dit, certaines croyances pour lesquelles nous entretenons une plus grande confiance, c'est-à-dire nos jugements bien pesés, doivent néanmoins constituer des points d'ancrage lors de l'élaboration ou de la construction d'une théorie. Abandonnant la prétention de rechercher des vérités fondamentales et optant plutôt pour une quête de cohérence comme source de justification, la méthode de l'équilibre réfléchi est parfois associée au holisme méthodologique développé entre autres par Willard van Quine et Nelson Goodman, deux collègues de Rawls à Harvard. Cela étant dit, bien qu'il existe des parallèles évidents entre ces approches, le concept rawlsien d'équilibre réfléchi s'applique exclusivement à la justification de jugements et de principes moraux tandis que le holisme de Quine et Goodman constitue une théorie épistémologique concernant la connaissance en général. Malgré toutes les affinités qu'on peut trouver entre les deux visions, il vaut donc mieux éviter de les confondre en une seule et apprécier les particularités de l'application de la méthode de l'équilibre réfléchi en philosophie morale.

Dans *Éthique économique et sociale*, Philippe van Parijs et Christian Aspersperger résumant ainsi la démarche préconisée par Rawls :

Elle consiste à tenter de formuler des principes généraux quant à ce que nous devons faire, individuellement ou collectivement, et à confronter les implications de ces principes à nos jugements moraux bien réfléchis, dans des circonstances (réelles ou hypothétiques) les plus diverses. Au cas où apparaît un conflit avec un jugement suffisamment ferme pour que nous ne soyons pas prêts à y renoncer, il nous faut rejeter le principe que nous avons cru pouvoir formuler, ou à tout le moins le réviser de manière à éliminer le conflit en question. [...] En cas de conflit, le principe invoqué doit être abandonné ou amendé jusqu'à ce que la cohérence soit restaurée. (Arnsperger et van Parijs, 2003, p. 9)

Ainsi, aucun principe n'est, au sens strict du terme, fondamental au sein d'une théorie éthique élaborée à partir de la méthode de l'équilibre réfléchi :

Une telle démarche ne prétend pas fournir aux questions éthiques une forme de fondement absolu. [...] Ce qui meut la démarche éthique est alors le souci de *cohérence*, le souci de formuler explicitement, et de la manière la plus simple possible, un ensemble de principes qui confèrent une unité à l'ensemble de nos jugements moraux face aux circonstances les plus diverses. (Arnsperger et van Parijs, 2003, p. 9-10)

Par ailleurs, en optant pour une démarche constructiviste qu'il qualifie lui-même de kantienne, Rawls souhaite distinguer clairement sa méthode de réflexion philosophique de celle de l'intuitionnisme rationnel¹ :

Une conception de la justice ne peut être déduite de prémisses ou de conditions préalables évidentes, s'appliquant aux principes eux-mêmes; au contraire, sa justification vient de ce que de multiples points de vue s'y retrouvent renforcés, de ce que tous les éléments s'y ajustent en une seule vision cohérente. (Rawls, 1987, p. 48)

[...] la justification repose sur la conception dans son ensemble et sur son accord avec nos jugements bien pesés mis en équilibre réfléchi ainsi que la façon dont elle les organise. Comme nous l'avons déjà noté, la justification est une question de soutien mutuel entre diverses considérations, d'accord de tous les éléments dans une doctrine cohérente. (Rawls, 1987, p. 620)

Rawls distingue en outre un équilibre réfléchi étroit d'un équilibre réfléchi large. Lorsque que nos convictions morales bien pesées sont toutes en accord avec les principes qu'une théorie éthique énonce, nous avons réussi à atteindre un équilibre réfléchi étroit. L'atteinte d'un équilibre réfléchi large implique que nous ayons également considéré toutes les autres alternatives théoriques connues et pertinentes et que la théorie qui est choisie est celle qui, après réflexion, articule de la façon la plus satisfaisante nos jugements bien considérés en un tout cohérent et intelligible. Bien entendu, on ne peut jamais être certain d'avoir examiné toutes les possibilités théoriques et il est donc possible qu'une théorie jugée satisfaisante à une époque soit éventuellement laissée de côté au profit d'une nouvelle théorie, améliorée. Il s'agit d'une humilité à laquelle nous contrainst le scepticisme de Hume dans toute démarche scientifique et avec laquelle tout philosophe doit savoir composer. Ainsi, dans *Théorie de la justice*, Rawls confronte les principes de sa théorie de la justice comme équité à celui des autres principales conceptions concurrentes, soit « *des variantes bien connues de l'intuitionnisme, du perfectionnisme et de l'utilitarisme* » (Rawls, 1987, p. 49). Cela étant dit, Rawls concentre son attention sur la doctrine utilitariste, théorie dominante en philosophie morale depuis environ deux siècles.

¹ De manière générale, l'intuitionnisme rationnel affirme qu'il existe « *une famille irréductible de principes premiers que nous devons mettre en balance les uns par rapport aux autres en nous demandant, par un jugement mûrement réfléchi, quel équilibre est le plus juste.* » (Rawls, 1987, p. 59) Ainsi, l'intuitionnisme rationnel nie la possibilité d'établir un critère constructif qui rendrait une conception de la justice cohérente et unifiée.

D'autre part, il est important de noter que, contrairement à l'utilitarisme, la théorie de la justice comme équité de John Rawls n'est pas une doctrine morale «compréhensive», c'est-à-dire qu'elle n'est pas une théorie générale englobante qui prétend donner une réponse à l'ensemble des questions morales. La théorie de Rawls a une portée plus limitée et se contente de déterminer quels principes régulant le fonctionnement des institutions sociales d'une société démocratique libérale contemporaine sont justes. Elle est ainsi mieux définie en tant que conception politique de la justice devant faire l'objet d'un consensus par recoupement dans un contexte pluraliste. Une bonne partie des écrits de Rawls ultérieurs à *Théorie de la justice*, dont l'article au titre fort évocateur *La théorie de la justice comme équité : une théorie politique et non pas métaphysique*, ont été consacrés à clarifier ce point et expliciter les idées qui y sont reliées. En limitant le champ d'application de sa théorie à un domaine et un contexte bien particulier et en déplaçant le centre d'intérêt de ses recherches de l'élaboration d'une théorie idéale vers une description de la pratique de la démocratie dans les sociétés libérales contemporaines, Rawls a donné l'impression à de nombreux observateurs qu'il avait changé de cap quant à l'orientation qu'il souhaitait voir prendre sa théorie. Ainsi, bien qu'il soit toujours demeuré fidèle à la très grande majorité des idées principales présentées dans *Théorie de la justice*, l'auteur de *Libéralisme politique* et de *The Law of Peoples* est fréquemment qualifié de deuxième Rawls, à distinguer du premier, auteur de *Théorie de la Justice*. Or, si certains ont pu être déçus de voir Rawls développer ses idées dans une direction qui ne leur plaît pas, on peut difficilement condamner ce dernier pour avoir voulu préciser sa pensée en reformulant sa théorie afin de répondre aux interprétations de sa théorie qu'il jugeait erronées et encore moins pour avoir accepté, suites aux innombrables critiques qui lui ont été faites, de réviser certaines de ses positions, en accord avec la méthode de l'équilibre réfléchi qu'il préconise.

1.3 Justification publique, pluralisme raisonnable et consensus par recoupement

La théorie de la justice comme équité de John Rawls constitue « *une conception philosophique destinée à une démocratie constitutionnelle* » (Rawls, 1987, p. 10). Selon Rawls, le type de justification que requiert une telle conception politique de la justice est une

justification publique. Par cela, il entend que non seulement les principes de justice qu'elle établit doivent être sujets à l'approbation des membres de la société, mais également que les raisons pour lesquelles ces principes sont choisis doivent être acceptables pour tous. Or, le défi est de taille étant donné le fait suivant :

Une société démocratique moderne est caractérisée non seulement par une pluralité de doctrines compréhensives, morales, philosophiques et religieuses, mais aussi par le fait que ces doctrines sont incompatibles entre elles tout en étant raisonnables. (Rawls, 1995, p. 4)

Comme l'indique Rawls, ce fait du pluralisme ne représente toutefois pas un désastre, mais est plutôt « *le résultat naturel de l'activité de la raison humaine quand elle se déroule dans le contexte d'institutions libres* ». (Rawls, 1995, p. 13) Le projet de *Libéralisme politique* est de démontrer comment le libéralisme politique parvient à surmonter la difficulté issue du fait du pluralisme raisonnable. C'est alors que prend forme le concept de consensus par recoupement. Puisque la co-existence d'une pluralité de conceptions du bien constitue « *un trait permanent de la culture publique de la démocratie* » (Rawls, 1995, p. 63), il vaut mieux renoncer à essayer de trouver un accord unanime sur des questions fondamentales d'ordre métaphysique et religieux. Il faut plutôt tenter d'élaborer une conception politique de la justice qui soit acceptable du point de vue de chacune des doctrines compréhensives raisonnables présentes. Il convient de préciser qu'une doctrine est dite raisonnable tant qu'elle « *ne rejette pas les éléments essentiels d'un régime démocratique* ». (Rawls, 1995, p. 4)

L'espoir qu'un tel consensus par recoupement soit possible découle de l'hypothèse que les membres des sociétés démocratiques contemporaines, s'ils sont raisonnables, partagent un socle de valeurs libérales communes gravitant autour du principe de tolérance. Ainsi, il leur est possible de s'entendre sur une conception politique de la justice au sein de laquelle se recoupent toutes les doctrines compréhensives raisonnables sans renoncer à leurs convictions plus profondes, s'ils en ont. Selon Rawls, la théorie de la justice comme équité représente justement une conception politique de la justice susceptible de faire l'objet d'un consensus par recoupement dans une société libérale contemporaine. De plus, Rawls est d'avis que sa théorie peut faire l'objet d'un accord « *suffisamment profond pour atteindre des idées comme celle de la société comme système équitable de coopération et des citoyens comme êtres libres et égaux, rationnels et raisonnables* ». (Rawls, 1995, p. 188) Ces idées,

qui reflètent le contractualisme de Rawls, peuvent être qualifiées d'idées principales sous-jacentes à la théorie de la justice comme équité. Elles font l'objet de la prochaine section.

1.4 Le contractualisme de Rawls

Le but principal de *Théorie de la justice* est de proposer une solution de rechange à la doctrine utilitariste, dominante en philosophie morale. En effet, malgré tout son succès, la tradition utilitariste apparaît insatisfaisante sur plusieurs points aux yeux de Rawls; certaines des implications de l'utilitarisme entrent en conflit avec ce qu'il estime être nos jugements bien pesés sur la justice sociale. Plus particulièrement, la logique agrégative qu'implique l'utilitarisme ne rend pas compte adéquatement du caractère individuel et séparé des personnes, ce qui amène Rawls à conclure que « *la pluralité des personnes n'est donc pas vraiment prise au sérieux par l'utilitarisme* » (Rawls, 1987, p. 53). Il souhaite mettre de l'avant, grâce à la théorie de la justice comme équité, « *une analyse systématique de la justice supérieure [...] à la tradition utilitariste* ». (Rawls, 1987, p. 20). Pour ce faire, Rawls opte pour une approche explicitement contractualiste. Sa stratégie consiste à tenter « *de généraliser et de porter à un plus haut degré d'abstraction la théorie traditionnelle du contrat social telle qu'elle se trouve chez Locke, Rousseau et Kant* ». (Rawls, 1987, p. 20). Ce faisant, il espère actualiser et revivifier une tradition philosophique ayant joué un rôle important dans l'histoire des démocraties modernes, notamment dans la rédaction de la Déclaration d'Indépendance américaine en 1776 et de la Déclaration des Droits de l'homme et du citoyen en France en 1789, mais qui semblait avoir été largement délaissée depuis.

Par ailleurs, il importe de préciser que l'idée du contrat social que Rawls présente dans *Théorie de la Justice* s'applique exclusivement à une société conçue « *comme un système clos, isolé des autres sociétés* » (Rawls, 1987, p. 34)². Une société, nous dit-il, se

² L'objet de *The Law of Peoples* est justement d'examiner comment la conception contractualiste de la justice comme équité peut s'étendre au droit international public. Dans cet ouvrage, Rawls se positionne de façon fort différente des penseurs cosmopolites, dont Thomas Pogge et Charles Beitz, qui avaient précédemment tenté d'élargir le champ d'application de l'idée du contrat social de l'échelle

défini comme étant « *une association, plus ou moins auto-suffisante, de personnes qui, dans leurs relations réciproques, reconnaissent certaines règles de conduite comme obligatoires, et qui, pour la plupart, agissent en conformité avec elles* ». (Rawls, 1987, p. 30) De plus, « *ces règles déterminent un système de coopération visant à favoriser le bien de ses membres* » (Rawls, 1987, p. 30). Une société représente ainsi une entreprise de coopération dont l'objectif est l'avantage mutuel des participants. Les termes fondamentaux de l'association qui les unit seront considérés justes si et seulement s'ils sont définis par des principes « *que des personnes libres et rationnelles, désireuses de favoriser leurs propres intérêts, et placés dans une position initiale d'égalité, accepteraient* ». (Rawls, 1987, p. 37) Cette idée d'accord originel entre êtres libres, égaux et rationnels représente ce qu'il convient d'appeler la principale idée contractualiste qui guide la théorie de la justice comme équité de John Rawls.

Notons d'autre part que le contrat social dont il est ici question est tacite, hypothétique et non-historique; personne ne signe réellement de contrat ni ne négocie effectivement un accord. Comme nous l'explique Rawls, « *le contrat est remplacé par une situation initiale qui comporte certaines contraintes dans la procédure de l'argumentation devant mener à l'accord originel sur les principes de justice* » (Rawls, 1987, p. 29). Cette situation initiale, c'est la fameuse position originelle, sur laquelle je reviendrai plus en détails d'ici peu. Pour l'instant, gardons simplement en tête que cet accord originel entre les membres coopérants d'une communauté porte sur les termes fondamentaux de leur association :

Pour nous, l'objet premier de la justice, c'est la structure de base de la société ou, plus exactement, la façon dont les institutions sociales les plus importantes répartissent les droits et les devoirs fondamentaux et déterminent la répartition des avantages tirés de la coopération sociale. Par institutions les plus importantes, j'entends la constitution politique et les principales structures socio-économiques. (Rawls, 1987, p. 33)

Ainsi, c'est sur les principes qui guideront la répartition des biens sociaux premiers par le biais de la structure de base de la société que porte l'entente.

nationale à l'échelle internationale. Il y affirme notamment que les principes de justice qu'il a élaborés pour une société close ne sauraient s'appliquer directement aux relations entre les peuples.

De façon générale, Rawls définit les biens premiers comme étant « *tout ce qu'on suppose qu'un être rationnel désire, peu importe ses autres désirs* » (Rawls, 1987, p. 122). Comme il nous l'explique, « *ces biens, normalement, sont utiles, quel que soit notre projet de vie* » (Rawls, 1987, p. 93). La liste des biens sociaux premiers qu'établit Rawls dans *Théorie de la Justice* comprend les éléments suivants : « *les droits, les libertés et les possibilités offertes, les revenus et la richesse* ». (Rawls, 1987, p. 123). À cela s'ajoute ce que Rawls identifie comme « *le bien premier peut-être le plus important* » (Rawls, 1987, p. 479) : le respect ou l'estime de soi-même. Il est en outre important de distinguer ces biens sociaux premiers d'autres biens premiers, naturels ceux-là. En effet, « *la santé et la vigueur, l'intelligence et l'imagination* » (Rawls, 1987, p. 93) sont certes des éléments ayant un impact majeur sur la façon dont les différents individus peuvent tirer profit des fruits de la coopération sociale et, ainsi, ils exercent une influence importante sur le bien-être et le développement de chacun. Cependant, la répartition de ces biens est le fruit de la « loterie naturelle » et échappe au contexte de la justice défini par Rawls. Cette répartition peut donc être heureuse ou malheureuse pour certains mais n'est toutefois ni juste ni injuste; elle constitue simplement un fait naturel contingent.

Avant de passer à la prochaine section, qui traite des circonstances de la justice, je souhaite identifier deux caractéristiques essentielles de l'accord originel. Premièrement, l'accord doit être tel qu'aucun être rationnel doté d'un sens de la justice ne pourrait raisonnablement refuser d'y adhérer. Autrement dit, les principes de justice qui y seront énoncés doivent être ceux que choisirait toute personne rationnelle placée dans des conditions raisonnables. Je reviendrai plus abondamment sur ce point lorsque je traiterai de la position originelle plus en détails et quand j'examinerai le raisonnement des partenaires qui s'y retrouvent. Deuxièmement, l'accord, une fois conclu, lie de manière définitive les participants : il n'est donc pas possible de s'entendre sur un accord que certains auraient ensuite l'intention de ne pas respecter. Ainsi, lorsque les partenaires trouvent un accord, ils souhaitent tous rationnellement que celui-ci soit respecté et, puisqu'ils sont raisonnables, ils comprennent que cela implique qu'ils doivent eux-même s'engager à le respecter. Ces deux caractéristiques expriment ce que Rawls considère comme les deux capacités fondamentales

des êtres rationnels et moraux, soit celles de se former une conception du bien et de développer un sens de la justice.

1.5 Les circonstances de la justice

Selon Rawls, un certain nombre de conditions d'arrière-plan doivent être réunies pour que l'on puisse traiter de justice sociale. Ces conditions sont ce que Rawls désigne, s'inspirant de Hume, comme étant les circonstances de la justice :

Les circonstances de la justice peuvent être définies comme l'ensemble des conditions normales qui rendent à la fois possibles et nécessaire la coopération humaine. C'est pourquoi [...], bien qu'une société soit une entreprise de coopération en vue d'un profit mutuel, elle se caractérise à la fois par un conflit d'intérêts et par une identité d'intérêts. Il y a identité d'intérêts puisque la coopération sociale procure à chacun une vie meilleure pour tous que celle que chacun aurait eue en cherchant à vivre seulement grâce à ses propres efforts. Il y a conflit d'intérêts puisque les hommes ne sont pas indifférents à la façon dont sont répartis les fruits de leur collaboration; en effet, dans la poursuite de leurs objectifs, ils préfèrent tous une plus grande part à une plus petite. On a donc besoin de principes pour choisir entre les différentes organisations sociales qui déterminent cette division des avantages et pour conclure un accord sur une répartition correcte. (Rawls, 1987, p. 159)

Dans la théorie de la justice comme équité de John Rawls, le concept de justice n'a de sens que dans ce contexte social bien précis. En effet, en l'absence des circonstances de la justice, il n'y a aucune place pour la vertu de la justice telle que la conçoit Rawls.

Par ailleurs, les circonstances de la justice se divisent en deux catégories : les circonstances objectives et les circonstances subjectives. Le contexte objectif de la justice part du fait qu'un grand nombre d'individus partage un espace géographique fixe et restreint. De plus, les individus ayant tous des capacités physiques et mentales semblables³, aucun ne peut prétendre dominer seul tous les autres et chacun est vulnérable aux attaques des autres, d'autant plus si leurs forces sont réunies. Enfin, les ressources se retrouvent dans une situation de rareté relative ; « *les ressources naturelles et autres ne sont pas abondantes au*

³ Rawls exclut de son analyse le cas des malades et des handicapés pour se concentrer sur des individus ayant des facultés "normales" leur permettant de participer activement à l'entreprise de coopération que représente la société.

point de rendre les systèmes de coopération superflus, mais les conditions ne sont pas non plus difficiles au point que des tentatives positives soient condamnées à l'échec ». (Rawls, 1987, p. 160) Ces circonstances objectives qui rendent la coopération humaine possible et nécessaire ne sont pas sans rappeler les conditions menant au pacte hobbesien, dont l'objectif est de mettre un terme à la guerre de chacun contre chacun dans l'état de nature.

Le contexte subjectif est quant à lui lié directement aux caractéristiques des partenaires, dont les similitudes et une certaine complémentarité au niveau de leurs intérêts rendent possible une coopération. Ce contexte, c'est aussi et surtout celui du pluralisme raisonnable, dont j'ai déjà traité précédemment. Comme je l'ai alors souligné, plusieurs conceptions raisonnables mais opposées du bien coexistent dans une société démocratique moderne. Ceci a pour conséquence que les membres d'une telle société poursuivent différents buts et objectifs qui entrent éventuellement en conflit. Or, ce conflit ne saurait être résolu à l'aide d'une conception de la justice qui engloberait toutes les conceptions du bien puisque celles-ci sont incompatibles en plusieurs points. D'où la pertinence et l'intérêt d'élaborer une conception politique et non métaphysique de la justice en mesure de susciter l'adhésion de tous à un consensus par recoupement quant aux principes de justice devant guider le fonctionnement des institutions sociales de base.

1.6 Les deux principes de justice

La théorie de la justice comme équité de John Rawls aboutit à deux principes bien connus. Or, ces deux principes représentent le cas particulier d'une conception plus générale de la justice. Voici comment Rawls exprime cette conception générale :

Toutes les valeurs sociales – liberté et possibilités offertes à l'individu, revenus et richesse ainsi que les bases sociales du respect de soi-même – doivent être réparties également à moins qu'une répartition inégale de l'une ou de toutes ces valeurs ne soit à l'avantage de chacun. (Rawls, 1987, p. 93)

Cette conception générale de la justice reflète une présomption en faveur de l'égalitarisme. En effet, une situation juste part d'un point de stricte égalité, duquel on accepte de s'écarter

uniquement si cela bénéficie à chaque personne. La formulation explicite des deux principes, et en particulier du second, détermine avec plus de précision dans quelles circonstances et sous quelles conditions peuvent être justifiées les écarts avec le point de stricte égalité.

Présentés pour la première fois au grand public dans *Théorie de la justice*, les deux principes de justice de John Rawls ont depuis fait l'objet d'un nombre impressionnant de commentaires et de critiques. Si bien que, même si l'essentiel du contenu de ces deux principes de justice est demeuré intact, Rawls lui-même fut amené à les reformuler à quelques reprises afin de préciser sa pensée et corriger certaines erreurs. On retrouve la version la plus achevée de deux principes de justice de Rawls dans son ouvrage le plus récent, *La justice comme équité : une reformulation de Théorie de la justice* (Rawls, 2004, p. 69-70)⁴ :

- (1) Chaque personne a une même prétention indéfectible à un système pleinement adéquat de libertés de base égales, qui soit compatible avec le même système de libertés pour tous.
- (2) Les inégalités économiques et sociales doivent remplir deux conditions : (a) elles doivent d'abord être attachées à des positions ouvertes à tous dans des conditions d'égalité équitable des chances : (b) ensuite, elles doivent procurer le plus grand bénéfice aux membres les plus défavorisés de la société (principe de différence).

Il importe par ailleurs de préciser que ces principes doivent être appliqués en suivant un ordre lexicographique de priorité absolue. Ainsi, le premier principe (1) a toujours préséance sur le second (2) et la première condition (a) du deuxième principe prime sur la seconde (b).

Le premier principe est aussi appelé « principe de la plus grande liberté égale pour tous ». Bien qu'il ait fait l'objet de quelques critiques⁵, ce principe est généralement accepté

⁴ La formulation « définitive » des deux principes présentée dans l'édition française de *Théorie de la Justice* (légèrement différente de celle originalement présentée dans la première édition anglaise) allait comme suit :

Premier principe : Chaque personne doit avoir un droit égal au système le plus étendu de libertés de bases égales pour tous qui soit compatible avec le même système pour les autres.

Second principe : Les inégalités sociales et économiques doivent être telles qu'elles soient : (a) au plus grand bénéfice des plus désavantagés, dans la limite d'un juste principe d'épargne, et (b) attachées à des fonctions et à des positions ouvertes à tous conformément au principe de la juste (fair) égalité des chances. (Rawls, 1987, p. 341)

⁵ Voir entre autres la critique de H.L.A. Hart dans *Rawls on Liberty and its Priority* (Hart, 1973).

sans problème au sein de la communauté des penseurs libéraux. Le second principe, plus controversé, contient, en fait, deux sous-principes complémentaires : le « principe de la juste (fair) égalité des chances » et le « principe de différence ». Le principe de différence, qui attribue au groupe le plus défavorisé une forme de droit de veto quant à tout écart du point d'égalité dans la répartition des avantages socio-économiques, est sans contredit l'un des éléments de la doctrine de Rawls qui a suscité le plus d'attention et engendré le plus de controverse. D'ailleurs, dans la préface de l'édition française de *Théorie de la Justice*, Rawls admet ceci :

Je continue à penser que le principe de différence est important et je continuerai à le défendre [...]. Mais il vaut mieux reconnaître que cette argumentation (*en faveur du principe de différence*) ne va pas de soi et n'aura jamais la force de celle en faveur des deux principes (lexicalement) antérieurs. (Rawls, 1987, p. 13)

Dans *Libéralisme politique*, Rawls va même jusqu'à affirmer que, contrairement au premier principe (lié aux libertés de base), les principes qui régissent les inégalités économiques et sociales (le principe de juste égalité des chances et le principe de différence) ne font pas partie des « *questions constitutionnelles fondamentales* ». (Rawls, 1995, p. 279) Il reconnaît également que « *quoiqu'un minimum social couvrant les besoins fondamentaux de tous les citoyens fasse également partie des questions constitutionnelles essentielles, [...] le « principe de différence » est plus exigeant et n'en fait pas partie.* » (Rawls, 1995, p. 277-278). À la lumière de tout ceci, il semble évident que Rawls est conscient de la faiblesse relative de ses arguments en faveur du second principe. Je reviendrai sur ces arguments d'ici peu. Pour l'instant, le temps est venu de faire quelques remarques sur la position originelle, contexte dans lequel les principes de justice sont choisis par les représentants des membres coopérants d'une société démocratique.

1.7 La position originelle

Rawls croit que ses principes de justice sont ceux que choisiraient des personnes rationnelles placées derrière un voile d'ignorance dans une position originelle hypothétique et devant s'entendre sur la répartition des bénéfices de leur coopération sociale. Comme je l'ai

déjà spécifié, sa théorie, suivant notamment Kant, se veut contractualiste : la justification des principes qu'elle met de l'avant se fonde sur le concept de contrat social. Or, l'argumentation à partir de la position originelle, bien que centrale dans *Théorie de la Justice*, ne constitue qu'un des trois grands pans du contractualisme dont se réclame Rawls. En effet, dans l'élaboration de sa théorie, Rawls considère non seulement les arguments issus de la position originelle mais également ceux liés à l'application des principes de justice aux institutions sociales d'une part et ceux relatifs à la stabilité de la justice comme équité d'autre part. Il ne s'agit donc pas simplement de déduire les principes de justice à partir d'une situation hypothétique prédéfinie (la position originelle), mais plutôt de mettre en équilibre nos convictions morales bien pesées sur les questions de justice sociale et de les articuler au sein de la construction théorique que constitue la position originelle. La formulation de l'argumentation à partir de la position originelle dans *Théorie de la justice* pouvait raisonnablement laisser croire au lecteur que le projet de Rawls était justement d'utiliser un raisonnement strictement déductif à partir de la position originelle pour justifier le choix des deux principes. Toutefois, les écrits ultérieurs de Rawls spécifiant la forme et le rôle de la méthode constructiviste qu'il emploie montrent qu'il s'agissait là d'une erreur d'interprétation⁶. Pour cette raison, je soutiens que la position originelle n'a pas un rôle fondateur dans la théorie de la justice comme équité mais a plutôt un rôle principalement heuristique.

Nous avons vu précédemment que l'accord originel doit être tel qu'aucun être rationnel doté d'un sens de la justice ne pourrait raisonnablement refuser d'y adhérer. Pour aboutir aux deux principes qui viennent d'être énoncés, Rawls modèle soigneusement les circonstances dans lesquelles le choix de ces personnes rationnelles s'effectue. L'argumentation de Rawls à partir de la position originelle repose sur une conception procédurale de la justice car, au niveau de la justification, ce qui importe d'abord et avant tout, c'est la procédure par laquelle les principes sont choisis et non les principes eux-mêmes. La position originelle est ainsi construite de façon à garantir l'équité entre des partenaires qui doivent s'entendre sur les termes de leur coopération future. Puisque Rawls soutient une

⁶ Voir notamment *Le constructivisme kantien dans la théorie morale* (in Rawls, 1993) et *Libéralisme politique* (Rawls, 1995).

conception de la justice qui est dite procédurale pure⁷, l'équité des circonstances dans lesquelles les principes de justice sont choisis se transmet ensuite directement aux principes eux-mêmes, d'où le concept de justice comme équité. Ainsi, les principes de justice choisis suite à une procédure équitable seront nécessairement équitables eux aussi. La construction de la position originelle est donc cruciale dans la justification complète des principes de justice mis de l'avant par Rawls car ses caractéristiques déterminent le choix des partenaires et assurent le caractère juste et équitable de ce choix.

Bien que les conditions générales de la vie sociale et les lois de la psychologie humaine soient connues de tous dans la position originelle, nul ne sait quelle position il occupe dans la société ni quels sont les traits particuliers de sa personnalité (ses talents, aptitudes et préférences personnelles). Ainsi, comme nous l'explique Rawls, « *chacun ignore sa propre conception du bien, les particularités de son projet rationnel de vie, ou même les traits particuliers de sa psychologie* ». (Rawls, 1987. p. 169) De plus, bien qu'ils connaissent les principes des théories économique et politique, les partenaires ne disposent d'aucune information sur le contexte particulier de leur société si ce n'est qu'elle est soumise aux circonstances de la justice. Tous ces éléments font partie du fameux voile d'ignorance derrière lequel sont placés les partenaires dans la position originelle afin d'éviter que toute information particulière, contingente et arbitraire n'influence leur choix.

Par ailleurs, les partenaires dans la position originelle sont pleinement et également rationnels. Ils souhaitent tous favoriser leur propre intérêt et sont mutuellement désintéressés dans le sens où ils ne soucient pas directement du sort des autres. Ainsi, même si l'idée sous-jacente est celle du contrat social, la symétrie de la situation des partenaires dans la position originelle a pour conséquence que le choix des principes de justice n'est pas le résultat d'une

⁷ Une justice procédurale pure se distingue d'une justice procédurale parfaite ou imparfaite. Une procédure parfaite implique que le résultat de la procédure est, de manière certaine, conforme au résultat souhaité et prédéterminé selon des critères extérieurs à la procédure elle-même. (Par exemple, si l'on souhaite partager un gâteau en parts égales, il suffit d'exiger que la dernière personne qui choisira son morceau soit celle qui coupe le gâteau.) Dans le cas d'une procédure imparfaite, il n'est pas certain que le résultat est toujours conforme à ce qui était souhaité mais ce devrait être le cas la plupart du temps. (Le système judiciaire est un exemple de procédure imparfaite.) Une procédure pure ne part pas quant à elle avec l'idée d'un résultat souhaité prédéterminé; c'est la procédure elle-même qui détermine et définit le résultat « juste ». (Voir Rawls, 1987, pp. 116-117)

négociation menant à un accord mais plutôt d'une procédure de décision imputée à un seul agent représentatif de l'ensemble des partenaires. En effet, les différents partenaires ayant accès à la même information en vertu du voile d'ignorance et étant dotés des mêmes capacités mentales, le raisonnement de chacun d'entre eux est identique à celui de tous les autres. Je reviendrai sur ce point dans le troisième chapitre. J'expliquerai alors comment la stratégie argumentative de Rawls pose quelques problèmes pour rendre compte de son engagement à l'égard d'une philosophie contractualiste. Pour l'instant, poursuivons le raisonnement effectué par les partenaires dans la position originelle.

Rawls soutient que dans de telles circonstances, étant donné l'importance de ce qui est en jeu, le principe utilitariste sera rejeté (aussi bien dans sa version classique, selon laquelle l'utilité totale des membres de la société doit être maximisée, que dans sa version plus moderne, qui veut que l'on maximise le niveau moyen d'utilité), et les partenaires préféreront des principes qui prennent davantage au sérieux le caractère individuel des intérêts de chacun. Ainsi, les partenaires voudront d'abord se doter d'un système de droits et libertés le plus étendu possible et ensuite s'assurer que les inégalités socio-économiques ne seront tolérées que dans la mesure où elles bénéficient au groupe le plus défavorisé de la société et à la condition que les différentes positions sociales soient ouvertes à tous tout en respectant une égalité équitable de chances, d'où le choix des deux principes sus-mentionnés. L'application de la théorie du choix rationnel dans ce contexte bien précis de la position originelle mène donc directement, selon Rawls, à la conclusion que les deux principes de justice qu'il propose seraient préférés au principe utilitariste. Je ne compte pas le faire ici, mais on peut remettre en question les conditions imposées par Rawls pour encadrer le choix des partenaires. En effet, il est possible que la position originelle qu'il décrit ne constitue pas le contexte idéal pour déterminer le caractère juste ou moral d'une entente; autrement dit, il se peut que Rawls n'ait pas réussi à atteindre le point archimédien qu'il recherche pour mettre en équilibre nos jugements bien pesés sur la justice sociale. Je préfère maintenant concentrer mon attention sur le raisonnement des partenaires, qui fait l'objet de la prochaine section consacré à deux concepts importants dans la théorie de Rawls : le rationnel et le raisonnable.

1.8 Le rationnel et le raisonnable

Dans la théorie de la justice comme équité, le rationnel est associé à la capacité des individus de se former une conception particulière du bien et de poursuivre les fins qui lui sont reliées par ce qu'ils estiment être les meilleurs moyens. Dans *Théorie de la Justice*, Rawls présente l'hypothèse de rationalité des partenaires de manière tout à fait conventionnelle lorsqu'il affirme :

L'hypothèse que les partenaires sont rationnels et mutuellement désintéressés revient alors à ceci : les personnes placées dans la position originelle essaient de reconnaître des principes qui favorisent autant que possible leurs systèmes de fins. (Rawls, 1987, p.176)

Il s'agit donc d'une rationalité instrumentale presque identique à celle qui est imputée aux agents dans la plupart des modèles élaborés par les économistes. Soulignons par ailleurs que Rawls ajoute une hypothèse particulière concernant la rationalité des partenaires, celle que les personnes rationnelles ne souffrent généralement pas d'envie. (Rawls, 1987, p. 175)

Le concept de rationalité occupe une place importante dans l'élaboration et la justification des deux principes de justice, bien qu'il soit exagéré de prétendre qu'il a un rôle fondateur. Dans *Théorie de la justice*, Rawls va toutefois jusqu'à avancer l'idée que sa théorie de la justice est une composante de la théorie du choix rationnel : « *La théorie de la justice est une partie, peut-être même la plus importante, de la théorie du choix rationnel* » (Rawls, 1987, p. 43). Plus tard, suite à certaines critiques, Rawls est amené à réviser cette affirmation. Ainsi, il corrige dans *La justice comme équité : une théorie politique et non pas métaphysique* ce qu'il considère désormais être une erreur :

C'était donc une erreur (et une source de graves malentendus) que de décrire la théorie de la justice comme une partie de la théorie du choix rationnel (voir *Théorie de la Justice*, p. 43 et 624). Ce que j'aurai dû dire, c'est que la théorie de la justice comme équité utilise une analyse du choix rationnel, mais soumise à des conditions raisonnables, pour décrire les délibérations des partenaires, représentatifs de personnes libres et égales, - tout cela dans le cadre d'une conception politique de la justice qui est aussi, bien entendu, une conception morale. (Rawls, 1993, pp. 223-224)

En effet, l'application de la théorie du choix rationnel dans la théorie de la justice de Rawls est limitée au contexte précis qui caractérise le moment fictif où se rencontrent les partenaires

pour établir les bases de leur coopération sociale, i.e. la position originelle. De plus, comme je l'ai déjà souligné, la position originelle est elle-même construite de façon à incorporer un certain nombre de convictions bien pesées et largement partagées sur des questions de justice sociale. Les conditions dans lesquelles le choix des partenaires s'effectue représentent ce que Rawls appelle le raisonnable. C'est de cette façon que Rawls soumet le rationnel aux contraintes du raisonnable dans la position originelle. Ainsi, dans la théorie de la justice comme équité, le raisonnable encadre le rationnel, ce qui reflète la priorité du juste sur le bien. Cette idée de la priorité du juste sur le bien est indispensable pour bien saisir une des critiques les plus importantes que fait Rawls à l'égard de l'utilitarisme. J'y reviendrai dans le troisième chapitre.

Contrairement au concept de rationnel, Rawls définit de manière relativement floue le concept de raisonnable, associé à la capacité des individus à développer un sens de la justice. En fait, la distinction entre le rationnel et le raisonnable n'est pas faite de manière explicite dans *Théorie de la Justice*. Dans *Libéralisme politique*, Rawls différencie plus nettement les deux concepts en invoquant la distinction kantienne entre impératifs catégoriques (commandés par le raisonnable) et hypothétiques (commandés par le rationnel) : « *Le premier (le raisonnable) représente la raison pratique pure, le second (le rationnel) la raison pratique empirique* ». (Rawls, 1995, p. 76) Pour donner un sens plus restreint au raisonnable dans le cadre d'une conception politique de la justice, Rawls s'appuie sur le texte de W.M. Sibley *The Rational versus the Reasonable* (Sibley, 1943). Essentiellement, sont raisonnables les gens qui « *sont disposés à diriger leur conduite par un principe à partir duquel eux-mêmes et les autres peuvent raisonner en commun* ». (Rawls, 1995, p. 77) On entre ainsi dans le champ de la raison publique, véritable domaine du raisonnable. À noter également : « *La disposition à être raisonnable n'est ni dérivée du rationnel ni opposée à lui.* » (Rawls, 1995, p. 77) Contrairement à Hobbes ou plus récemment à David Gauthier⁸, le projet de Rawls n'est donc pas de développer une théorie contractualiste de la justice exclusivement fondée sur la motivation individuelle et l'intérêt privé. Pour Rawls, les concepts de raisonnable et de

⁸ Voir *Morale et Contrat* (Gauthier, 2000)

rationnel apparaissent comme étant indépendants mais complémentaires⁹. On peut en outre dire que, dans la théorie de la justice comme équité de John Rawls, non seulement le raisonnable conditionne-t-il le rationnel en lui imposant ses contraintes, mais en même temps, il le présuppose puisqu'il implique que les individus sont déjà dotés d'une conception du bien. En effet, même si, dans la position originelle, les partenaires ignorent quelle est leur conception du bien, ils sont conscients qu'ils en ont une et reconnaissent l'importance que celle-ci aura sur leur projet rationnel de vie.

1.9 Les arguments en faveur de la règle du maximin

Je souhaite maintenant examiner plus en détails les arguments en faveur la règle de décision qui amène, selon Rawls, les partenaires dans la position originelle à choisir le principe de différence plutôt que le principe d'utilité¹⁰. En effet, le choix du principe de différence, qui stipule que l'on doit améliorer autant que possible le sort du groupe le plus défavorisés de la société, implique que, dans la position originelle, les partenaires souhaitent maximiser la quantité minimale de biens premiers qu'ils peuvent obtenir (d'où l'expression «maximin»). Le choix du principe d'utilité moyenne impliquerait pour sa part que les partenaires voudraient plutôt maximiser l'espérance d'utilité découlant des différentes quantités de biens sociaux premiers qu'ils peuvent obtenir et de la probabilité qu'ils estiment de se retrouver dans chacune des positions sociales associées à ces différentes quantités. Tout d'abord, notons que les deux règles, celle du maximin et celle de la maximisation de l'utilité espérée, peuvent aboutir au même résultat, c'est-à-dire au choix des deux principes de justice rawlsiens. Pour cela, il faut que les individus concernés manifestent une aversion infinie pour le risque, c'est-à-dire que, de manière névrotique, ils craignent toujours que leurs choix entraînent les pires conséquences pour eux-même et sont ainsi littéralement paralysés par la

⁹ Pour une discussion critique et éclairante sur le sujet, voir le texte de Geneviève Nootens *La nature de la complémentarité entre le raisonnable et le rationnel chez Rawls* (Nootens, 1996).

¹⁰ Comme le souligne Rawls dans la préface de l'édition française de *Théorie de la justice*, la plus sérieuse option de rechange utilitariste à ses deux principes de justice est représentée par une conception mixte dans laquelle « le principe d'utilité (moyenne) remplace dans le deuxième principe le principe de différence ». (Rawls, 1987, p. 12)

peur du danger que cela représente à leurs yeux. Ne concentrant leur attention que sur le pire résultat envisageable, de telles personnes n'accordent pas la moindre importance aux autres résultats possibles, pas plus qu'aux probabilités associées aux diverses possibilités. Or, Rawls prend bien soin de préciser que le choix des deux principes dans la position originelle ne repose pas sur une attitude particulière face au risque. En fait, comme nous l'avons vu, le voile d'ignorance exclut toute information relative aux traits psychologiques particuliers des partenaires, y compris leur attitude vis-à-vis du risque.

Par ailleurs, l'épaisseur du voile d'ignorance écarte toute connaissance relative à la façon dont la population est distribuée dans les divers groupes sociaux. Cela évite la possibilité que les partenaires, estimant plus probable de se retrouver parmi les groupes les plus nombreux, favorisent injustement ceux-ci uniquement en raison de leur taille, et ce au détriment des groupes minoritaires. De plus, Rawls exclut la mise en application du principe de raison insuffisante par les partenaires. Selon ce principe bayésien, en l'absence d'information pertinente objective sur la distribution des probabilités entre n événements possibles, on attribue la même probabilité ($1/n$) à chaque événement. Ainsi, les individus placés derrière un voile d'ignorance qui appliqueraient le principe de raison insuffisante, ne sachant pas quelle position sociale ils occupent ni comment les différentes positions sont distribuées statistiquement, estimerait que la probabilité d'occuper chacune des positions possibles est égale. Or, nous dit Rawls, il n'y a aucune place pour des argumentations probabilistes dans le contexte de la position originelle. Selon lui, une évaluation des probabilités reposant sur des bases purement subjectives n'a aucun sens. Puisque aucune information objective ne permet d'évaluer la probabilité d'occuper une position sociale ou une autre, les partenaires ne font aucune évaluation de ce genre. C'est pourquoi ils concentrent leur attention exclusivement sur la pire position possible.

L'impossibilité pour les partenaires d'estimer d'une manière ou d'une autre la probabilité de se retrouver dans l'une ou l'autre des différentes positions sociales envisageables les empêche de considérer la situation de la position originelle comme une forme de pari. C'est exactement ce que souhaite Rawls en excluant carrément toute forme de raisonnement probabiliste de la procédure de choix des partenaires. En effet, si on les laissait

procéder à une évaluation subjective des probabilités, les partenaires pourraient éventuellement préférer prendre le risque de se retrouver dans une situation plus défavorable que la pire position possible en vertu du principe de différence si cela permettait d'améliorer les perspectives des autres positions envisageables. Autrement dit, maximiser le gain minimum ne serait pas nécessairement le choix qu'effectueraient des êtres rationnels utilisant un raisonnement probabiliste. Or, c'est justement ce possible arbitrage entre risque et gain espéré que Rawls veut à tout prix éviter. En fait, malgré tous les éléments de la construction de la position originelle qui pointent en ce sens, il refuse de voir le choix des partenaires comme un problème de maximisation sous contraintes. Cela lui permet non seulement d'élaborer le principe de différence mais également d'établir la priorité absolue du premier principe de justice sur le second¹¹. En effet, plutôt que de se garantir de manière définitive des droits et libertés de base aussi étendues que possible, les partenaires pourraient préférer prendre le risque de sacrifier certains de ces biens sociaux pour des avantages économiques potentiellement plus intéressants. Cela implique d'une part que, selon Rawls, aucun compromis impliquant des libertés de base moins étendues au profit de plus grands avantages socio-économiques ne peut être effectué¹². Cela signifie d'autre part que toute détérioration de la situation du groupe le plus défavorisé est inacceptable, peu importe les conséquences que cette détérioration pourrait impliquer pour toutes les autres positions. Rawls prétend ainsi mettre de l'avant une caractéristique essentielle de la justice sociale, à savoir qu'il est toujours injuste de sacrifier le sort des plus défavorisés, même s'ils sont excessivement peu nombreux, au profit des autres groupes, peu importe l'ampleur des avantages que cela peut

¹¹ Comme je l'ai dit précédemment, l'argumentation à partir de la position originelle ne constitue qu'un des trois grands pans de la défense contractualiste des deux principes de justice par Rawls. Je ne considère ici que les arguments issus du contexte de l'accord originel, ce qui implique que bon nombre d'autres arguments présentés par Rawls en faveur du principe de différence et de la priorité absolue du premier principe sur le second ne sont nullement affectés par ce que j'avance.

¹² En fait, Rawls admet, dans sa conception générale de la justice, que dans des conditions non-favorables, il peut être rationnel de sacrifier certains droits et certaines libertés fondamentales au profit de plus grands avantages socio-économiques. Toutefois, lorsque les conditions favorables sont réunies, la conception spéciale s'applique et tout arbitrage entre avantages socio-économiques et droits et libertés fondamentales s'avère irrationnel. Cette indétermination qui découle de l'ambiguïté entre la conception générale et la conception spéciale de la justice amène David Lyons à se questionner sur le fait qu'il y ait finalement une aussi grande différence que le prétend Rawls entre l'utilitarisme et sa théorie de la justice comme équité, en particulier lorsque vient le temps de l'appliquer à des cas concrets. (Lyons, 1972) Je traiterai plus en détails de cette question dans le troisième chapitre.

procurer à un grand nombre d'individus et celle des désavantages imposés aux plus défavorisés.

Rawls semble convaincu de la justesse de son intuition sur ce point mais il est tout de même permis de la remettre question. Empirer légèrement la situation de quelques personnes si cela améliore grandement le sort de plusieurs autres est-il nécessairement injuste, peu importe l'ampleur des conséquences et le nombre de personnes impliquées? Permettez-moi d'en douter en toute sincérité. Plus essentiellement pour le point qui nous intéresse ici, des êtres rationnels placés en situation d'incertitude préféreront-ils toujours se prémunir contre le risque de se retrouver dans la pire position envisageable sans tenir compte des perspectives associées aux autres positions? Certainement pas; cela reviendrait à faire preuve d'une prudence excessive, voire irrationnelle. Rawls lui-même concède ce point lorsqu'il affirme que « *la règle du maximin n'a jamais été proposée comme un principe de décision rationnelle dans tous les cas de risque et d'incertitude* » (Rawls, 2004, p. 139). Rawls maintient toutefois sa position quant à la pertinence de l'application de la règle du maximin au contexte bien particulier du choix dans la position originelle. Or, pour ce faire, il se doit d'exclure tout argument probabiliste de son analyse, dénaturant par le fait même le concept de choix rationnel individuel en situation d'incertitude. Cela simplifie certes le problème auquel les partenaires sont confrontés, mais à quel prix? Peut-on maintenir une conception conventionnelle de la rationalité tout en affirmant que l'évaluation des probabilités ne peut faire partie du raisonnement d'un être rationnel devant effectuer un choix alors qu'il est incertain du résultat que son choix entraînera comme conséquence pour lui-même? Je soutiendrai que non. La conséquence pour la théorie de Rawls en est que ses arguments contre l'utilitarisme tirés de la position originelle et faisant usage de la théorie du choix rationnel en situation d'incertitude s'avèrent moins clairs et moins convaincants que des arguments plus directs en faveur des deux principes de justice qu'il élabore.

Une opposition fondamentale entre Rawls et l'utilitarisme, peu importe dans quelle version, concerne la pertinence morale de procéder à une agrégation quand vient le temps de déterminer ce qui constitue l'intérêt public. Alors que l'utilitariste doit nécessairement additionner les utilités individuelles, pondérées ou non, à un moment ou un autre dans son

calcul de l'utilité sociale, Rawls refuse de procéder à une telle sommation de la quantité de biens sociaux premiers qu'obtiennent les différents individus. Ce faisant, Rawls prend davantage au sérieux le caractère séparé et individuel des personnes que ne le fait l'utilitarisme. Plus encore, en s'assurant que les inégalités seront toujours à l'avantage des plus défavorisés, Rawls prétend que personne n'aura de raison de se sentir injustement désavantagé par le fonctionnement des institutions de base de la société. La stabilité que cela génère constitue selon Rawls un avantage marqué de sa théorie par rapport à la doctrine utilitariste, dont le poids des sacrifices qu'elle exige des individus pour promouvoir le bien commun représente une source d'instabilité potentielle importante. Je reviendrai sur ces questions importantes dans le troisième chapitre.

1.10 Conclusion

L'étendue de l'œuvre de John Rawls est telle que le bref résumé de sa théorie de la justice comme équité que je viens d'exposer ne fait qu'effleurer quelques points pourtant essentiels. Par ailleurs, j'ai dû négliger certains aspects pour me concentrer sur les éléments que je juge les plus importants pour comprendre l'utilisation que fait Rawls de la théorie du choix rationnel dans le contexte particulier de la position originelle. Cela ne veut pas dire que ce dont je n'ai pas ou peu parlé est sans intérêt, loin de là. Seulement, j'ai dû me contenter de traiter les questions reliées à la problématique choisie.

Je propose par ailleurs de retenir deux éléments principaux de la théorie de Rawls pour la suite de ce mémoire. Premièrement, nous garderons en tête que Rawls affirme que l'on peut découvrir ses principes de justice à partir de la position originelle en s'imaginant ce que serait le choix rationnel d'un individu ignorant la position sociale qu'il occupe. Dans le prochain chapitre, j'exposerai l'argumentation de John Harsanyi, qui prétend que c'est plutôt un principe utilitariste qui serait choisi par un être rationnel placé dans un tel contexte d'incertitude. Deuxièmement, nous tâcherons de ne pas perdre de vue que Rawls est à la recherche d'une justification publique de type contractualiste et que sa conception de la justice doit faire l'objet d'un consensus par recoupement au sein d'une société pluraliste.

Dans le troisième chapitre, je soutiendrai qu'il existe une certaine tension entre ce second élément et le premier. Cela dit, je tenterai également de montrer comment Rawls parvient tant bien que mal à surmonter cette difficulté. Je soutiendrai alors que la stratégie argumentative de Rawls n'était peut-être pas la meilleure, tout compte fait.

CHAPITRE II

LA THÉORIE MORALE UTILITARISTE DE JOHN HARSANYI

2.1 Introduction

Dans *Théorie de la Justice*, John Rawls cible la doctrine utilitariste comme la principale rivale de sa propre théorie. Il convient de se demander dans un premier temps vers quelle(s) version(s) de l'utilitarisme les attaques de Rawls étaient dirigées. Il dit estimer que les écrits de Sidgwick représentent « *la formulation la plus claire et la plus accessible* » (Rawls, 1987, p. 49) de la stricte doctrine classique. Comme l'indique Rawls, « *l'idée principale en est qu'une société est bien ordonnée et, par là même, juste, quand ses institutions majeures sont organisées de manière à réaliser la plus grande somme totale de satisfaction pour l'ensemble des individus qui en font partie* » (Rawls, 1987, p. 49). Rawls identifie par ailleurs deux principes utilitaristes distincts. Il y a tout d'abord le principe d'utilité classique, qui exige la maximisation de l'utilité sociale totale comprise comme la simple addition arithmétique des utilités individuelles. Le principe d'utilité moyenne exige quant à lui la maximisation de l'utilité moyenne des membres de la société. Comme le souligne Rawls, les deux principes coïncident si la population d'une société reste exactement stable. (Rawls, 1987, p. 193) Toutefois, la maximisation de l'utilité totale peut impliquer une diminution de l'utilité moyenne parallèlement à un accroissement excessif de la population. Du point de vue de la position originelle, la supériorité du principe d'utilité moyenne paraît ainsi évidente. C'est donc le principe d'utilité moyenne qui représente aux yeux de Rawls le concurrent le plus sérieux à ses deux principes de justice. (Rawls, 1987, p. 194)

L'utilitarisme possède de multiples facettes, se présente sous différentes versions et a connu récemment d'importants développements théoriques. On peut donc raisonnablement se

demander comment les critiques que Rawls adresse à l'utilitarisme sont affectées par les travaux des utilitaristes contemporains. À la lumière des écrits subséquents de Rawls et suite au développement d'autres théories concurrentes comme le libéralisme et le communautarisme, il appert que la position vis-à-vis de l'utilitarisme qu'il exprimait dans certains passages de *Théorie de la Justice* doit être nuancée et replacée dans son contexte historique. En fait, c'est toute la position de Rawls au sujet de l'utilitarisme qui mérite d'être réexaminée et précisée. Il s'agit là d'un projet de grande envergure auquel je tenterai humblement de contribuer dans le chapitre qui suivra celui-ci. L'objectif principal du présent chapitre sera de présenter la version de l'utilitarisme que j'estime être la plus intéressante à comparer avec la théorie de Rawls, soit celle de l'économiste et philosophe américain d'origine hongroise John Harsanyi¹ (1920-2000).

Pour l'instant, il est donc préalablement nécessaire de traiter plus en détails de la doctrine utilitariste et plus particulièrement de la théorie morale utilitariste d'Harsanyi, qui considère l'éthique comme une branche de la théorie du comportement rationnel au service des intérêts communs de la société prise comme un tout. (Harsanyi, 1982, p. 43) De son aveu, les grands philosophes qui ont le plus influencé sa propre pensée sont les suivants : Adam Smith (1723-1790), pour avoir mis de l'avant l'idée du point de vue moral juste et équitable d'un observateur impartial et sympathique; Kant (1724-1804), pour sa distinction formelle entre les règles morales et les autres types de règles comportementales établies par le critère d'universalité; et bien entendu les penseurs de la tradition utilitariste, tels Bentham, Mill, Sidgwick et Edgeworth, qui ont fait de la maximisation de l'utilité sociale le critère moral de base. (Harsanyi, 1982, p. 40) Faisant un usage abondant de la théorie moderne du choix rationnel en situation de risque et d'incertitude, aussi appelée théorie bayésienne de la décision rationnelle, l'approche d'Harsanyi présente, de ses propres dires, « *une combinaison unique d'analyse philosophique et de raisonnement mathématique* ».² (Harsanyi, 1982, p. 44) Cette approche lui permet de défendre de manière originale le principe de la

¹ Harsanyi fut récipiendaire du Prix Nobel d'économie en 1994, conjointement avec John Nash et Reinhard Selten. La contribution importante de ces trois grands noms de la théorie des jeux à la science économique était ainsi reconnue.

² Traduction libre de : « *a unique combination of philosophical analysis and of mathematical reasoning* ».

maximisation de l'utilité moyenne à l'aide d'un modèle mathématique qui, comme je le montrerai, n'est pas dénué de considérations morales.

Pour bien introduire la théorie morale de John Harsanyi, je commencerai par retracer rapidement l'évolution historique du concept d'utilité depuis l'époque de Bentham jusqu'à nos jours, où il est au cœur des modèles économiques néo-classiques. Je montrerai alors comment l'abandon d'une conception cardinale de l'utilité au profit d'une conception purement ordinale a d'importantes implications théoriques pour l'utilitarisme en tant que théorie normative, notamment en ce qui a trait à la possibilité de procéder à des comparaisons interpersonnelles de l'utilité. J'expliquerai ensuite que la théorie morale d'Harsanyi repose non pas sur l'ancienne conception hédoniste de l'utilité mais plutôt sur la version plus moderne de l'utilitarisme qui s'exprime en termes de préférences. Je montrerai ensuite comment la théorie morale d'Harsanyi, par son recours aux fonctions d'utilité attendue de type von Neuman-Morgenstern, implique un retour à une certaine conception cardinale de l'utilité, beaucoup plus subtile que la conception classique mais permettant néanmoins de comparer d'un point de vue impartial les différentes utilités individuelles des membres d'une société. Je distinguerai par la suite l'utilitarisme de la règle dont se réclame Harsanyi de l'utilitarisme de l'acte. Puis, je présenterai le modèle développé par Harsanyi pour formaliser sa théorie morale : le modèle d'équiprobabilité pour les jugements moraux. Je traiterai finalement de la critique faite par Harsanyi à l'égard de la théorie de Rawls, soulignant au passage quelques similarités et différences importantes entre l'analyse d'Harsanyi et l'argumentation de Rawls à partir de la position originelle. Ceci nous mènera vers le troisième et dernier chapitre, dans lequel je tenterai de montrer comment l'utilitarisme peut fournir aux deux principes de justice de Rawls une justification différente mais complémentaire à celle mise de l'avant par l'auteur de *Théorie de la justice*.

2.2 L'évolution historique du concept d'utilité

Jeremy Bentham est généralement identifié comme le père fondateur de l'utilitarisme, doctrine morale conséquentialiste influente en philosophie politique. John

Stuart Mill et Henry Sidgwick, deux des disciples de Bentham, ont également contribué largement à la diffusion et la popularité de cette doctrine morale dans le monde anglo-saxon. Or, comme le souligne Francisco Vergara, citant notamment Épicure ainsi que les fondements du droit romain, l'idée fondamentale derrière le principe d'utilité est très ancienne. Ainsi, selon Vergara,

[...] la place éminente que Bentham occupe dans l'histoire de cette doctrine est surtout justifiée par l'enthousiasme qu'il communiqua aux penseurs de son époque, en avançant non seulement une méthode (une manière d'appliquer le principe) mais aussi une foule de propositions en vue de réformer la société anglaise. (Vergara, 2005, p. 2)

En effet, à l'époque de Bentham, le principe d'utilité commandait une réforme profonde et radicale quant à la manière dont les institutions publiques étaient gérées. Pour qu'elles soient considérées justes, nous dit Bentham, ces institutions doivent être organisées de manière à assurer le plus grand bonheur du plus grand nombre. Telle est la règle fondamentale de l'utilitarisme : les actions bonnes sont celles qui maximisent le bien-être agrégé.

Bentham concevait l'utilité comme étant une mesure cardinale du bien-être des individus, un indice de bonheur humain. Cela implique que l'on puisse chiffrer le niveau de bien-être qu'un individu atteint, de la même manière que l'on peut mesurer un poids en grammes ou une taille en centimètres, par exemple. Toutefois, encore aujourd'hui, aucun instrument de mesure n'a pu être développé pour déterminer objectivement et numériquement l'utilité ressentie par un être vivant. Étant donné l'extrême complexité de la psychologie humaine, on peut en outre douter qu'un tel instrument puisse un jour être mis au point. Ainsi, bien que l'on puisse aisément concevoir l'utilité d'un point de vue hédoniste comme étant simplement la différence entre les plaisirs et les peines ressenties³, cette ancienne version cardinale et hédoniste du concept d'utilité s'avère inapplicable d'un point de vue formel.

De nos jours, les économistes, pour la plupart, ne conçoivent plus l'utilité comme étant une mesure cardinale du bien-être des individus, mais plutôt comme une mesure ordinale associée à la satisfaction des préférences individuelles. En effet, selon cette nouvelle

³ Selon cette définition, une peine n'est ni plus ni moins que l'opposé d'un plaisir ou, autrement dit, un « plaisir négatif ».

interprétation, l'utilité ne sert qu'à établir un classement des différentes options selon un ordre de préférence. Le nombre associé à un niveau d'utilité ne veut rien dire en soi; il ne sert qu'à établir un rang entre les alternatives envisageables. De plus, toute notion hédoniste relative aux expériences sensorielles de plaisir et de peine est carrément élaguée de la théorie micro-économique du consommateur. En effet, la relation entre la satisfaction des désirs et la sensation de bien-être ne fait pas partie de l'analyse. On dit alors simplement qu'un individu qui possède un ensemble de préférences logiquement non-contradictoires et qui agit rationnellement s'efforce d'entreprendre les actions qui lui permettent de satisfaire au maximum ces préférences. L'ensemble de préférences d'un individu ne représente ainsi rien de plus que le classement subjectif qu'il fait des diverses possibilités qui s'offrent à lui. Bien entendu, ce classement dépend des conséquences qu'une personne envisage sur son bien-être. La théorie économique contemporaine ne dit toutefois rien, ou très peu, sur la manière dont se forment ces préférences et se garde bien de dire quoi que ce soit sur ce que leur satisfaction apporte au bonheur de l'Homme; pour l'économiste, les préférences des agents sont données, tout simplement.

Le passage d'une interprétation cardinale du concept d'utilité à une interprétation ordinale a d'importantes implications. Tout d'abord, tout ce qu'on peut désormais dire à propos du niveau d'utilité qu'un individu associe à une option, c'est qu'il est plus élevé ou moins élevé que le niveau d'utilité qu'il associe aux autres options; en aucun cas est-on autorisé à mesurer l'ampleur de la différence entre différents niveaux d'utilité atteints. De plus, toute comparaison interpersonnelle du niveau d'utilité atteint devient impossible. Il s'agit là d'un problème majeur pour l'utilitarisme : comment peut-on procéder à la sommation des utilités individuelles si l'unité de mesure utilisée n'a pas de valeur cardinale et si l'on ne peut rendre commensurables les niveaux d'utilité atteints par plusieurs personnes différentes? Cette difficulté aiguilla les économistes néo-classiques vers la révolution marginaliste, à laquelle on peut identifier entre autres les noms célèbres de Léon Walras, William Stanley Jevons, Carl Menger, Francis Ysidro Edgeworth et Wilfredo Pareto.

Placés devant l'impossibilité de procéder à l'agrégation des utilités individuelles, une vaste majorité d'économistes s'est en effet écarté de la tradition de Bentham. Renonçant à

l'idée de maximiser en tant que telle la somme totale des utilités individuelles, ces économistes ont ainsi dû revoir à la baisse la portée morale de leur théorie, se contentant d'affirmer que l'allocation des ressources « *qui maximise la somme des utilités est nécessairement telle que les utilités marginales de tous les individus sont égales.* » (Arnsperger et Van Parijs, 2003, p. 24) L'utilité marginale correspond à la variation d'utilité que procure à un individu la possession d'une petite quantité supplémentaire d'un bien. Tant et aussi longtemps que les utilités marginales de tous les individus ne sont pas égales, il est possible d'augmenter l'utilité totale en procédant à une réallocation des ressources. En effet, si l'utilité marginale que procure à un individu un peu plus d'un certain bien est supérieure à l'utilité marginale que perd un individu en étant dépossédé de la même quantité de ce bien, on peut augmenter la somme des utilités de ces deux personnes en procédant à une redistribution en faveur de la première. Par exemple, si le fait de disposer de 100 dollars de plus par mois procure davantage d'utilité à une mère monoparentale sans emploi qu'à un millionnaire, la maximisation de la somme des utilités implique que cet argent devrait être alloué à la mère de famille, dont l'utilité marginale est plus élevée. Si on fait par ailleurs l'hypothèse généralement plausible de l'utilité marginale décroissante, c'est-à-dire si l'on suppose que les premières unités d'un bien procurent habituellement plus d'utilité que les dernières, l'utilitarisme présente alors un fort penchant en faveur de l'égalitarisme.

Cela dit, le critère utilitariste qui émane de la révolution marginaliste n'est nul autre que le critère d'efficacité de Pareto : « *une situation A est Pareto-optimale s'il n'existe aucune autre situation possible B telle que, d'une part, un individu au moins préfère strictement B à A, tandis que, d'autre part, aucun individu ne préfère A à B.* » (Arnsperger et Van Parijs, 2003, p. 24) Autrement dit, tant et aussi longtemps qu'il est possible d'améliorer le sort d'un individu sans détériorer celui d'aucun autre, on doit le faire. De toute évidence, cette conclusion représente une assertion morale plutôt faible. Il ne fait aucun doute que si une personne peut accroître son bien-être sans que cela nuise à qui que ce soit d'autre, aucune raison (morale ou autre) ne peut justifier qu'on l'en empêche. Le critère de Pareto est un critère d'unanimité; il garantit que personne ne verra sa situation empirer par un changement dans la répartition des ressources rares. Cependant, devant plusieurs distributions Pareto-optimales possibles, on ne dispose d'aucun critère pour déterminer laquelle est la plus juste

ou la plus équitable. Ainsi, une répartition fortement inégalitaire des ressources qui concentre une large part de ces ressources dans les mains de quelques individus peut être efficace au sens de Pareto au même titre qu'une répartition plus égalitaire qui permet à tous de mener une vie décente. Il s'agit là d'une limite évidente du point de vue d'une théorie normative; étant d'abord et avant tout un critère d'efficacité économique qui garantit que tous les gains d'utilité résultant d'échanges libres sont exploités et non un principe de justice distributive, le critère de Pareto s'avère insuffisant aux yeux de la plupart des philosophes.

Le concept d'utilité marginale revêt par ailleurs d'intéressants avantages pour les économistes. Il leur permet d'élaborer des modèles formels en utilisant les outils mathématiques du calcul différentiel, donnant à l'analyse économique une forme nettement plus scientifique. Il y a toutefois un fort prix à cela : « *une contraction drastique des questions traitables par la maxime utilitariste.* » (Arnsperger et Van Parijs, 2003, p. 23) Par ailleurs, les modèles des économistes reposent sur un certain nombre d'hypothèses sur les préférences individuelles. Ces hypothèses concernent essentiellement la cohérence interne des préférences et assurent que toute contradiction logique est exclue de la relation de préférence. Ainsi, on suppose que la relation de préférence est une relation complète et continue (elle couvre toutes les options et combinaisons possibles), réflexive (A est au moins aussi préférable que A) et transitive (si A est préféré à B et B est préféré à C, alors A est préféré à C). En ajoutant quelques hypothèses sur le fonctionnement des marchés⁴, la nouvelle économie du bien-être (welfare economics) parvient à formuler le théorème suivant :

Le Premier Théorème de l'Économie du Bien-être garantit qu'un marché concurrentiel exploite tous les gains découlant de l'échange : une allocation d'équilibre réalisée grâce à un ensemble de marchés concurrentiels est nécessairement efficace au sens de Pareto. Une telle allocation peut très bien ne présenter aucune autre caractéristique désirable, mais elle est nécessairement efficace. Notons que ce théorème ne dit rien concernant la distribution des gains économiques.

⁴ On peut résumer ces hypothèses sur le fonctionnement des marchés ainsi : absence d'externalités et d'asymétries d'information dans un marché de concurrence pure et parfaite en équilibre. Pour de plus amples explications, voir n'importe quel manuel de microéconomie relativement récent (dont Varian, 1997, p. 559-560).

L'équilibre de marché pourrait très bien ne pas être une allocation « juste »⁵. (Varian, 1997, p. 554)

La théorie économique néo-classique parvient ainsi à justifier le système de marché concurrentiel, donnant une nouvelle dimension à la métaphore de la main invisible d'Adam Smith. En effet, elle démontre mathématiquement que la combinaison des actions d'une multitude de personnes poursuivant leur intérêt personnel tend à mener vers un équilibre efficace socialement, quoique que possiblement injuste.

On peut bien entendu remettre en question le caractère vraisemblable de certaines hypothèses de base des modèles économiques. En particulier, les hypothèses relatives aux préférences individuelles peuvent être critiquées sous prétexte qu'elles ne sont pas conformes aux comportements que l'on observe dans le monde réel. De plus, une véritable situation de concurrence pure et parfaite ne s'est jamais vue, diront certains. On peut également soutenir à juste titre que les marchés ne sont jamais en équilibre simultané; ils s'ajustent constamment mais ne permettent jamais d'atteindre l'optimum de Pareto, même s'ils y tendent. Toutes ces questions sont certes fort intéressantes mais elles sont toutefois d'ordre empirique et n'enlèvent rien à l'élégance de la construction théorique édifiée par les économistes, qui repose sur des hypothèses idéalisées. En ce sens, l'économie du bien-être est bel et bien normative et non pas descriptive; elle décrit ce qui serait idéalement ou ce qui devrait être et non ce qui est effectivement. D'un point de vue de la théorie morale, il s'agit ensuite de déterminer si l'idéal évoqué correspond à nos aspirations ou si celles-ci vont au-delà du seul souci d'efficacité. À ce sujet, les implications d'un deuxième théorème de l'économie du bien-être sont fort intéressantes :

Le Second Théorème de l'Économie du Bien-être stipule que sous certaines conditions, toute allocation efficace au sens de Pareto peut être réalisée par un équilibre concurrentiel. Quelle est la portée ce théorème? Il implique en fait que les problèmes de distribution et d'efficacité peuvent être séparés. Quelle que soit l'allocation efficace au sens de Pareto que vous désirez, elle peut être obtenue par l'intermédiaire d'un système de marché. Celui-ci est neutre d'un point de vue distributif; quelle que soit votre définition d'une « bonne » ou d'une « juste »

⁵ L'utilisation des guillemets témoigne bien, à mon avis, du malaise qu'éprouvent aujourd'hui bon nombre d'économistes à s'aventurer sur un terrain qui semble étranger à leur analyse en traitant de questions de justice sociale.

distribution du bien-être, vous pouvez utiliser les marchés concurrentiels pour la réaliser. (Varian, 1997, p. 554)

Une redistribution « juste » des dotations initiales ou, autrement dit, des ressources rares (naturelles et sociales), juxtaposée à un système de marché concurrentiel permet de satisfaire à la fois les impératifs de justice, d'une part, et d'efficacité, d'autre part. Il reste alors à déterminer ce qu'est une distribution initiale « juste » des ressources rares, question à laquelle la plupart des économistes contemporains préfèrent ne pas répondre.

Par ailleurs, on est en droit de se demander si le concept d'utilité auquel font aujourd'hui référence les économistes néo-classiques correspond toujours au concept mis de l'avant par Bentham, Mill et Sidgwick. De toute évidence, un glissement conceptuel majeur s'est effectué avec l'abandon de la cardinalité de l'utilité. Alors que l'utilité était pour les utilitaristes classiques directement liée aux sensations de plaisir et de peine, nous avons vu que ce n'est plus le cas avec l'approche contemporaine des préférences. D'ailleurs, Wilfredo Pareto préféra utiliser le terme « ophélimité », moralement neutre, pour décrire la mesure de la satisfaction des préférences des agents économiques par l'entremise du marché, distinguant ainsi le concept de celui d'utilité⁶. Cette terminologie n'a toutefois pas réussi à s'imposer et on emploie souvent indistinctement le mot utilité, peu importe qu'on fasse référence à une mesure du bonheur (concept associé à l'utilitarisme classique de Bentham) ou à une mesure de la satisfaction des préférences (concept présent dans les modèles économiques contemporains). Comme le souligne Francisco Vergara, bonheur et satisfaction ne sont cependant pas nécessairement synonymes dans l'esprit des utilitaristes. (Vergara, 2005, p. 4) Le bien-être économique, notion technique que l'on peut définir clairement et formellement, est sans doute une composante importante du bonheur humain. Mais ce dernier est une notion immensément plus complexe et ne saurait se résumer à la satisfaction maximale des préférences des agents économiques. De manière plus générale, une théorie morale sérieuse ne saurait se contenter de dire que le bonheur est atteint par la satisfaction des désirs, la question pertinente devenant alors de déterminer quels désirs doivent être satisfaits. J'expliquerai maintenant, dans la prochaine section, que la théorie morale de John Harsanyi

⁶ L'expression « désirabilité » a également été suggérée, notamment par Alfred Marshall. La forte connotation morale qu'elle possède n'en fait toutefois pas un meilleur candidat.

se veut un utilitarisme dont l'objet est la satisfaction des préférences éthiques individuelles et comment il parvient à exclure les préférences irrationnelles et antisociales de son analyse.

2.3 L'utilitarisme expérientialiste et l'utilitarisme préférentialiste

La théorie morale d'Harsanyi est basée sur un modèle individualiste des jugements moraux. Selon lui, les jugements de valeur concernant le bien-être d'une société constituent ni plus ni moins qu'une classe spéciale de jugements de préférences exprimés par des individus moraux. (Harsanyi, 1953, p. 434) La particularité des jugements éthiques est qu'ils se doivent d'être impartiaux et impersonnels. Ainsi, Harsanyi distingue les préférences subjectives d'un individu, c'est-à-dire les préférences qu'il manifeste sur la base de son intérêt personnel, de ses préférences éthiques, qui représentent ce qu'il préférerait s'il basait son jugement uniquement sur des considérations sociales impersonnelles. Autrement dit, les préférences éthiques d'un individu représentent « *ce qu'il préfère dans ces moments possiblement rares où il s'efforce d'adopter une attitude impartiale et impersonnelle spéciale.* »⁷ (Harsanyi, 1955, p. 315) Dans l'esprit d'Harsanyi, chaque personne possède donc deux ensembles de préférences bien distincts. D'un côté, tout individu a des préférences personnelles, qui guident son comportement de tous les jours et qui peuvent s'exprimer sous la forme de fonctions d'utilité individuelle. Ces préférences personnelles manifestent généralement et naturellement une certaine partialité en faveur des intérêts personnels de l'individu et de ceux des êtres qui lui sont chers. De l'autre côté, tout individu a également des préférences morales, qui peuvent ou non avoir une influence importante sur son comportement de tous les jours mais qui, contrairement à ses préférences personnelles, doivent donner un même poids égal aux intérêts de tous, y compris les siens. Ces préférences morales peuvent s'exprimer sous la forme de fonctions de bien-être social ou d'utilité sociale. (Harsanyi, 1982, p.47)

⁷ Traduction libre de : « *what he prefers in those possibly rare moments where he forces a special and impersonal attitude upon himself* ».

La théorie utilitariste d'Harsanyi définit l'utilité sociale à partir des utilités individuelles et l'utilité individuelle exprime des préférences personnelles. En bout de ligne, l'utilité sociale est donc définie en termes de préférences individuelles, ce qui fait que l'approche d'Harsanyi peut être qualifiée d'utilitarisme préférentialiste (preference utilitarianism). Il faut distinguer cet utilitarisme préférentialiste de l'utilitarisme expérientialiste, incarné tant par la version hédoniste de Bentham, qui définit l'utilité individuelle aussi bien que sociale en termes de sensations de plaisir et de peine, que par l'utilitarisme idéal de Moore, qui définit utilités individuelle et sociale en termes de quantités d'états mentaux de valeur intrinsèque. (Harsanyi, 1982, p. 54) Comme le souligne Harsanyi, « *l'utilitarisme des préférences est la seule forme d'utilitarisme compatible avec l'important principe philosophique d'autonomie des préférences.* »⁸ (Harsanyi, 1982, p. 55) Par cela, il entend que l'utilitarisme des préférences, contrairement à l'utilitarisme hédoniste ou l'utilitarisme idéal, respecte le principe selon lequel le critère ultime pour déterminer ce qui est bon ou mauvais pour une personne ne saurait être autre que ses propres désirs et préférences. Cela dit, les préférences auxquelles s'intéresse la théorie morale d'Harsanyi sont les préférences concernant la société qu'un individu manifesterait dans des circonstances idéales. Ainsi, ce qui est bon ou mauvais pour la société est défini en fonction du point de vue d'une seule personne, qui se doit cependant de manifester une parfaite impartialité dans son jugement.

Cela dit, Harsanyi propose une autre distinction importante, celle entre les préférences actuelles ou manifestes d'un individu et ses vraies préférences. En effet, l'éthique ne saurait servir à satisfaire des désirs ou des préférences frivoles ou carrément irrationnelles. Alors que les vraies préférences d'un individu sont celles qu'il manifesterait dans des conditions idéales, suite à une réflexion appropriée basée sur toutes les informations pertinentes, ses préférences actuelles explicites peuvent quant à elle résulter de croyances erronées, d'erreurs de logique ou d'émotions vives qui altèrent le jugement rationnel. (Harsanyi, 1958, p. 312) Ainsi, les préférences déclarées d'un individu ne concordent pas nécessairement avec ses réelles préférences, mais c'est néanmoins exclusivement à ces

⁸ Traduction libre de : « *preference utilitarianism is the only form of utilitarianism consistent with the important philosophical principle of preference autonomy* ».

dernières qu'une théorie morale utilitariste doit porter attention selon Harsanyi. (Harsanyi, 1982, p. 55) De plus, Harsanyi soutient que certaines préférences antisociales doivent être écartées, et ce même si elles peuvent constituer les vraies préférences d'un individu. C'est le cas des préférences reflétant « *le sadisme, l'envie, le ressentiment et la malice.* »⁹ (Harsanyi, 1982, p. 56) De telles préférences ne pourraient servir à construire une fonction d'utilité sociale adéquate puisqu'elles sont incompatibles avec l'objectif même de l'établissement de règles sociales, à savoir atteindre un plus haut niveau d'utilité grâce à une coopération mutuellement avantageuse que ces règles permettent et favorisent. Voyons maintenant plus en détails comment Harsanyi construit cette fonction d'utilité sociale à l'aide du concept d'utilité attendue et de la théorie bayésienne de la décision rationnelle.

2.4 Le modèle d'équiprobabilité pour les jugements moraux

La théorie utilitariste de John Harsanyi fait un usage crucial du concept d'utilité attendue, proposé par Bernouilly en 1738 mais repris et développé par les économistes John von Neumann et Oskar Morgenstern en 1944 dans leur désormais célèbre *Theory of Games and Economic Behavior*. Les fonctions d'utilité attendue de von Neumann et Morgenstern permettent une modélisation formelle du comportement rationnel des individus placés en situation de risque et d'incertitude. L'idée intuitive derrière le concept d'utilité attendue est la suivante : lorsque les conséquences d'un choix sont incertaines, un individu rationnel évalue les différentes options qui s'offrent à lui en fonction de leur probabilité de se réaliser et de leur impact éventuel sur le niveau d'utilité qu'il pense atteindre si elles se réalisent effectivement. De manière plus formelle, cela équivaut à dire qu'en situation de risque ou d'incertitude, un individu rationnel fait des choix basés sur l'espérance mathématique E des n utilités u qu'il associe aux différentes possibilités¹⁰ ayant chacune une probabilité correspondante p tel que :

⁹ Traduction libre de : « *sadism, envy, resentment and malice* ».

¹⁰ Les différentes possibilités doivent être conjointement exhaustives (de sorte qu'au moins une des possibilités évoquées se réalisera effectivement) et mutuellement exclusives (de sorte que pas plus d'une seule possibilité évoquée ne se réalisera effectivement).

$$U_A = E(U) = (u_1 \times p_1) + (u_2 \times p_2) + (u_3 \times p_3) + \dots + (u_n \times p_n)$$

Ce type de fonction d'utilité mesurant l'utilité attendue comme somme pondérée d'utilités est aussi appelé fonction de type von Neumann-Morgenstern et, en raison de la forme particulière qu'elle revêt ainsi que des propriétés qui y sont rattachées¹¹, permet d'intéressantes manipulations mathématiques pour les économistes. La théorie de l'utilité attendue permet entre autres aux modèles économiques d'expliquer le fonctionnement du marché de l'assurance et de fournir une interprétation rationnelle du comportement des parieurs. Par ailleurs, et c'est ce qui s'avère particulièrement intéressant du point de vue d'Harsanyi, la fonction d'utilité attendue d'un individu permet d'obtenir une mesure cardinale d'utilité qui représente l'importance relative qu'il accorde à différentes alternatives. (Harsanyi, 1982, p. 83)

Reprenant à sa façon l'idée de l'observateur impartial d'Adam Smith, Harsanyi soutient qu'un jugement de valeur sur une question sociale comme la distribution des revenus manifesterait une parfaite impartialité si la personne qui effectuait ce jugement était complètement ignorante de la position qu'elle occupe dans la société. Ce serait le cas, poursuit-il, si cette personne estimait qu'elle a exactement la même chance d'occuper chacune des positions possibles. (Harsanyi, 1953, p. 434) Ce postulat, appelé postulat d'équiprobabilité, permet à Harsanyi de recourir à la théorie bayésienne de la décision rationnelle d'une manière particulière pour construire son modèle de décision morale, appelé modèle d'équiprobabilité pour les jugements moraux. Le recours à la théorie bayésienne nécessite cependant l'ajout de deux axiomes supplémentaires aux deux axiomes de base permettant la dérivation du concept formel d'utilité en situation de certitude, ceux de complétude et de continuité de la relation de préférences. Le premier de ces deux axiomes additionnels, celui d'équivalence probabiliste, implique une indifférence complète vis-à-vis de la loterie elle-même ou, autrement dit, du processus aléatoire qui détermine quel résultat possible se réalisera. Bien entendu, une telle hypothèse ne permettrait pas de rendre compte

¹¹ La fonction d'utilité attendue a la particularité mathématique d'être unique à une transformation affine positive près. Comme l'explique Varian, « cela signifie simplement que vous pouvez lui appliquer une transformation affine; vous obtiendrez une autre fonction d'utilité attendue qui représente les mêmes préférences. » (Varian, 1997, p. 235)

adéquatement du comportement d'un joueur qui retire un plaisir du simple fait de prendre part à un jeu de hasard mais, selon Harsanyi, elle est tout à fait appropriée pour une théorie morale conséquentialiste étant donné « *l'obligation d'adopter une attitude purement orientée vers les résultats lorsque l'on prend des décisions morales importantes.* »¹² (Harsanyi, 1978, p. 225) Le second axiome additionnel est l'axiome d'indépendance des alternatives non-pertinentes, aussi appelé « sure thing principle » en anglais. Selon ce principe, « *les choix que les gens prévoient de faire dans un état de la nature devraient être indépendants des choix qu'ils prévoient de faire dans les autres états.* » (Varian, 1997, p. 237) Malgré toutes les critiques qui ont pu être formulées relativement à la vraisemblance de cette hypothèse, Harsanyi indique à juste titre que l'axiome d'indépendance peut toujours être satisfait si l'on procède à une définition adéquate des divers résultats possibles¹³. (Harsanyi, 1978, p. 224) La conclusion à retenir est la suivante : si les quatre axiomes bayésiens de la rationalité sont respectés au sein de l'ensemble des préférences d'un individu placé en situation d'incertitude, et ils le seront si ces préférences sont pleinement rationnelles et débarrassées de toute incohérence et contradiction logique, celles-ci peuvent être représentées par une fonction d'utilité attendue de type von Neumann-Morgenstern.

Suivant notamment Fleming¹⁴, Harsanyi définit l'utilité sociale totale W comme étant la somme pondérée par un facteur α des n utilités individuelles u , tel que :

$$W = (\alpha_1 \times u_1) + (\alpha_2 \times u_2) + (\alpha_3 \times u_3) + \dots + (\alpha_n \times u_n)$$

Le postulat d'équiprobabilité fait en sorte que des préférences dites morales, c'est-à-dire reflétant un point de vue impartial, attribuent le même poids à l'utilité de chaque individu, de sorte que :

¹² Traduction libre de : « *the obligation of taking a purely result-oriented attitude in making important moral decisions* ».

¹³ Par exemple, on ne manifestera vraisemblablement pas les mêmes préférences en cas de pluie ou de beau temps devant une « loterie » permettant d'obtenir un parapluie. Toutefois, une définition adéquate des « prix » associés à cette « loterie » se résumerait simplement à ceci : demeurer au sec ou se faire mouiller.

¹⁴ Voir FLEMING, M. (1952) « A Cardinal Concept of Welfare », *Quarterly Journal of Economics*, Vol. 66, p. 366-384

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_n = 1/n$$

Ainsi, la fonction d'utilité attendue U_A reflétant les préférences morales d'un observateur impartial revêt la forme générale suivante :

$$U_A = (1/n) \times (u_1 + u_2 + u_3 + \dots + u_n)$$

Cette équation n'est ni plus ni moins que celle qui permet de calculer l'utilité attendue moyenne et le modèle d'équiprobabilité pour les jugements moraux parvient ainsi à défendre le principe d'utilité moyenne. Plus exactement, Harsanyi affirme que pour satisfaire des préférences morales reflétant une parfaite impartialité, un individu rationnel doit maximiser l'utilité moyenne des personnes qu'il pourrait être.

On constate que l'approche d'Harsanyi implique une certaine conception cardinale du bien-être et donc, de son propre aveu, un retour inattendu à des idées de la période pré-parétienne. (Harsanyi, 1955, p. 309) Cela s'avère nécessaire car, comme il l'explique :

[...] aucune théorie éthique ou théorie de l'économie du bien-être satisfaisante n'est possible sans comparaisons interpersonnelles de l'utilité. D'un autre côté, de telles comparaisons n'ont aucune place dans des modèles basés uniquement sur la poursuite rationnelle de l'utilité individuelle.¹⁵ (Harsanyi, 1961, p. 183)

Or, le modèle d'équiprobabilité pour les jugements moraux va au-delà de la simple poursuite rationnelle de l'utilité individuelle puisque, comme nous l'avons vu, il impose une condition d'impartialité aux préférences morales et écarte les préférences irrationnelles et antisociales. D'autre part, Harsanyi observe ceci :

Dans la vie de tous les jours nous faisons, ou au moins essayons de faire, constamment des comparaisons interpersonnelles de l'utilité. [...] Je ne crois pas que ce soit la tâche du philosophe ou du spécialiste des sciences sociales de nier le fait évident que les gens se sentent souvent capables de faire de telles comparaisons. Plutôt, sa tâche est d'expliquer comment nous parvenons à faire de telles comparaisons – tant bien que mal.¹⁶ (Harsanyi, 1982, p. 49-50)

¹⁵ Traduction libre de : « *no satisfactory theory of ethics or welfare economics is possible without interpersonal comparisons of utility. On the other hand, such comparisons have no place in models based solely on the rational pursuit of individual utility* ».

¹⁶ Traduction libre de :

Cela étant dit, l'objet des préférences morales que définit Harsanyi est bien particulier. Il s'agit en effet non pas du choix des actions appropriées à entreprendre mais plutôt des règles morales les plus à même de favoriser l'intérêt commun des individus qui devront les respecter. La distinction entre utilitarisme de la règle et utilitarisme de l'acte fait l'objet de la prochaine section.

2.5 L'utilitarisme de l'acte et l'utilitarisme de la règle

L'utilitarisme de la règle (*rule utilitarianism*), version plus subtile et améliorée, se distingue de l'utilitarisme de l'acte (*act utilitarianism*) en affirmant que ce n'est pas l'utilité des actions individuelles prises isolément qui importe mais plutôt l'utilité engendrée par la mise en place et le respect de règles contraignant et encadrant ces actions. L'utilitarisme de la règle, dont on trouve des germes dans la défense humienne des droits individuels, fut plus récemment popularisé par Harsanyi, qui dit avoir été inspiré des travaux de l'économiste d'Oxford Roy Forbes Harrod et plus spécifiquement du texte de 1936 *Utilitarianism Revised*. Cela dit, la terminologie a été introduite plus tard par Richard Brandt. Toutefois, Harrod, nous dit Harsanyi, « a été le premier à faire ressortir les avantages de l'utilitarisme de la règle par rapport à l'utilitarisme de l'acte. »¹⁷ (Harsanyi, 1982, p. 41)

L'utilitarisme de la règle permet de rendre compte du fait, en apparence contradictoire et gênant pour les défenseurs de la pensée utilitariste, que le respect de certaines règles ou l'application de certains principes ou règles autres que le principe d'utilité au sein d'une société permet à cette dernière d'atteindre un niveau d'utilité plus élevé que l'application du principe utilitariste à toutes les situations prises isolément. (Harsanyi, 1985, p. 44) Autrement dit, en termes plus techniques, l'utilitarisme de la règle explique comment,

In everyday life, we make, or at least attempt to make, interpersonal utility comparisons all the time. [...] I do not think it is the task of the philosopher or a social scientist to deny the obvious fact that people often fell quite capable of making such comparisons. Rather, his task is to explain how we ever manage to make such comparisons – as well or as badly as we do make them.

¹⁷ Traduction libre de : « *was the first to point out the advantages of rule utilitarianism over act utilitarianism* ».

dans un contexte de coopération sociale, une maximisation sous contrainte permet d'obtenir un meilleur résultat en termes d'utilité qu'une maximisation sans contrainte. Bref, le fait d'établir certaines règles contraignantes génère un plus haut niveau d'utilité qu'en l'absence de telles règles. Du point de vue de l'utilitarisme de l'acte, une action individuelle moralement juste est une action qui, dans la situation où se retrouve un individu, maximise l'utilité sociale. La perspective de l'utilitarisme de la règle est différente :

[...] une action moralement juste est une action qui se conforme à la bonne règle morale applicable dans le type de situation où se retrouve un individu, alors qu'une bonne règle morale est cette règle comportementale particulière qui maximiserait l'utilité sociale si elle était suivie par tous dans toutes les situations sociales de ce type.¹⁸ (Harsanyi, 1982, p. 41)

Dans le cas de l'utilitarisme de l'acte, on peut ainsi parler d'utilitarisme direct tandis que l'utilitarisme de la règle est un utilitarisme qu'on peut qualifier d'indirect¹⁹.

Malgré cette distinction importante entre utilitarisme de l'acte et utilitarisme de la règle, l'idée fondamentale sous-jacente demeure la même : d'un point de vue moral, le meilleur choix est toujours celui qui permet d'atteindre le plus haut niveau d'utilité sociale. Ce qui change, c'est la perspective analytique à partir de laquelle on évalue la situation. Plutôt que de se demander quelles sont les actions qui maximisent l'utilité sociale, on s'interroge sur les règles ou les principes les plus à même de favoriser, en bout de ligne, les intérêts des membres de la société. On pourrait donc dire que l'utilitarisme de la règle, en tenant compte de la relation existante entre les actes isolés et les règles sociales établies qui était obnubilée dans les versions antérieures de l'utilitarisme, permet de raffiner la théorie utilitariste. Grâce à cela, celle-ci peut désormais réaffirmer plus solidement le principe selon lequel la meilleure action d'un point de vue moral est celle qui a pour conséquence de maximiser le niveau d'utilité sociale, *toutes choses bien considérées*. Il n'y a donc pas une

¹⁸ Traduction libre de : « *a morally right act is one that conforms with to the correct moral rule applicable to this sort of situation, whereas a correct moral rule is that particular behaviour rule that would maximise social utility if it were followed by everybody in all situations of this particular type* ».

¹⁹ Dans son livre *Political Philosophy*, Dudley Knowles propose une classification qui place l'utilitarisme de la règle aux côtés d'une autre forme d'utilitarisme indirect, qui lui est complémentaire et qui concentre quant à lui son attention sur les dispositions ou les traits de caractères des individus plutôt que directement sur les actions qu'ils effectuent. (Knowles, 2001, pp. 23-35)

opposition stricte entre utilitarisme de la règle et utilitarisme de l'acte mais plutôt un prolongement naturel, une continuité progressive du point de vue théorique. En ce sens, la classification de l'utilitarisme en termes d'utilitarisme de l'acte et d'utilitarisme de la règle peut paraître arbitraire²⁰ puisque même les utilitaristes classiques étaient conscients que l'important n'est pas tant les conséquences particulières d'une action mais plutôt les conséquences qu'elle tend généralement à produire. Cela étant dit, la nuance conceptuelle n'en demeure pas moins utile d'un point de vue heuristique.

Par ailleurs, Harsanyi soutient qu'il existe bel et bien une différence formelle importante entre l'utilitarisme de l'acte et l'utilitarisme de la règle puisque le problème de maximisation posé par les deux approches n'est pas logiquement équivalent. Les deux versions de l'utilitarisme pointent vers le même objectif, c'est-à-dire la maximisation de l'utilité sociale. Cependant, « *les deux théories imposent des contraintes mathématiques très différentes au problème de maximisation de l'utilité.* »²¹ (Harsanyi, 1982, p. 57) La différence découle du fait qu'un utilitariste de l'acte envisage les choses d'un point de vue statique tandis qu'un utilitariste de la règle le fait d'un point de vue dynamique. En effet, un utilitariste de l'acte choisit sa propre stratégie en considérant que les stratégies de tous les autres individus sont données et constantes. Un utilitariste de la règle, quant à lui, voit sa propre stratégie de même que celles de tous les autres utilitaristes de la règle comme des variables à être déterminées par le processus de maximisation. De plus, le critère d'universalité impose que la stratégie choisie par tous les utilitaristes de la règle soit identique; personne n'est autorisé à profiter du fait que les autres respectent une règle sans s'engager à la respecter soi-même. Comme le souligne Harsanyi, il résulte de ces différences formelles que « *l'utilitarisme de la règle est en bien meilleure position pour organiser des stratégies de coopération et de coordination entre différentes personnes.* »²² (Harsanyi, 1982, p. 57) Toutefois, poursuit Harsanyi, l'avantage principal de l'utilitarisme de la règle réside plutôt « *dans son habileté à expliquer les implications que des systèmes alternatifs de règles*

²⁰ C'est l'opinion de Francisco Vergara. Voir Vergara, 2005, p. 5.

²¹ Traduction libre de : « *the two theories impose very different mathematical constraints on this maximisation problem* ».

²² Traduction libre de : « *rule utilitarianism is in much better position to organize cooperation and strategy coordination among different people* ».

morales possibles auraient sur les attentes et les incitations des personnes. »²³ (Harsanyi, 1982, p. 58) Cela dit, la question de savoir si l'utilitarisme de la règle est en meilleure position que la doctrine du contractualiste de Rawls pour rendre compte des effets de coordination, d'attentes et d'incitations demeure ouverte. La complexité de cette question m'oblige toutefois à la mettre de côté pour l'instant. J'y reviendrai cependant dans le prochain chapitre. Je propose maintenant de traiter, dans la prochaine section, de l'objet du désaccord entre Rawls et Harsanyi, c'est-à-dire l'identification du principe devant déterminer le choix des partenaires dans la position originelle.

2.6 Le désaccord entre Rawls et Harsanyi

Le point de vue impartial qu'est forcé d'adopter l'utilitariste de la règle qui choisit un ensemble de règles morales dans le modèle d'équiprobabilité pour les jugements moraux d'Harsanyi n'est pas sans rappeler la position originelle dans laquelle se retrouvent les partenaires de la théorie de la justice comme équité de Rawls. En effet, dans les deux cas, les personnes devant faire un choix sont ignorantes de leur position sociale. Cela dit, l'épaisseur du voile d'ignorance de Rawls impose des conditions plus strictes quant aux informations pouvant être utilisées pour procéder au dit choix. Comme je l'ai souligné vers la fin du premier chapitre, Rawls exclut notamment tout recours par les partenaires à des raisonnements probabilistes. J'ai alors dit que Rawls, refusant que des probabilités subjectives ou logiques puissent servir de base au choix rationnel en l'absence de connaissance relative aux probabilités empiriques objectives, rejette le principe de raison insuffisante. Rawls se prive ainsi des outils de la théorie bayésienne de la décision rationnelle. Cela dit, Harsanyi ne fait pas pour sa part un usage injustifié du principe de raison insuffisante. Comme je l'ai montré dans la section précédente, Harsanyi défend en fait sur des bases morales, à travers l'idée de l'observateur impartial, son postulat d'équiprobabilité. Ce postulat lui permet ensuite, comme le fait le principe de raison insuffisante, d'attribuer une probabilité égale à la possibilité pour l'observateur impartial

²³ Traduction libre de : « *in its ability to take proper account of the implications that alternative systems of possible moral rules would have for people's expectations and incentives* ».

d'occuper chacune des différentes positions sociales. Or, ce point de vue de l'observateur impartial, propre à la tradition utilitariste, n'est pas celui qu'adopte Rawls pour construire sa théorie de la justice comme équité. En effet, la perspective de Rawls est plutôt celle du contrat social. Il s'agit là d'une différence importante, fondamentale même, entre l'approche de Rawls et celle d'Harsanyi.

Selon Harsanyi, le problème principal avec l'argumentation de Rawls à partir de la position originelle réside dans le fait que le choix des partenaires est basé sur une règle de décision hautement irrationnelle, à savoir la règle du maximin, et non sur la plus plausible maximisation de l'utilité attendue. (Harsanyi, 1982, p. 47) La règle du maximin viole l'axiome de continuité de la relation de préférences d'un individu rationnel. (Harsanyi, 1975, p. 595) Elle implique que la personne qui l'applique ne considère pas tous les résultats possibles et les probabilités qui y sont associées pour faire un arbitrage entre risque et gain espéré mais concentre son attention uniquement sur le pire résultat possible, sans tenir compte de la probabilité que ce résultat se réalise, et ce même s'il estime que cette probabilité est infime. Or, pour reprendre des termes fréquemment utilisés par les économistes, quelqu'un qui baserait toutes ses décisions sur la règle du maximin manifesterait une aversion absolue pour le risque ou, autrement dit, une peur infinie du danger. Une telle attitude est de toute évidence irrationnelle, et même carrément malade; dans la vie de tous les jours, elle mènerait à une inaction paralysante.

Rawls est bien conscient du problème soulevé par Harsanyi mais il soutient malgré tout que le choix des partenaires n'est pas basé sur une attitude déraisonnable ou irrationnelle vis-à-vis du risque. C'est plutôt l'épaisseur du voile d'ignorance et le caractère fondamental de ce qui est en jeu qui les amène à prendre une décision conforme à la règle du maximin. Rawls s'accorde avec Harsanyi lorsqu'il affirme : « *Il est clair que la règle du « maximin » n'est pas en général un guide qui convient pour des choix dans l'incertain. Elle vaut seulement dans des situations caractérisées par certains traits particuliers.* » (Rawls, 1987, p. 185) Rawls prétend que la position originelle possède justement ces caractéristiques qui sont les suivantes : elle fournit de bonnes raisons de ne pas tenir compte des probabilités; les gains au-dessus du minimum n'ont que peu d'intérêt; et les options rejetées ont des

conséquences pratiquement inacceptables. (Rawls, 1987, p. 185-186) J'ai déjà discuté abondamment de la question des probabilités du point de vue de la théorie bayésienne de la décision rationnelle. Il semble que les bonnes raisons auxquelles fait référence Rawls doivent avoir d'autres bases que la théorie du choix rationnel en situation d'incertitude, ce qui conforte l'idée que les véritables arguments de Rawls contre le principe d'utilité peuvent être avancés indépendamment de la construction de la position originelle. Quant aux deux autres caractéristiques mentionnées, elles décrivent des situations précises dans lesquelles l'application de la règle du maximin équivaut grosso modo à celle de la maximisation de l'utilité. En effet, si les gains au-dessus du minimum n'ont que peu d'intérêt, cela revient à dire que leur utilité marginale est quasi nulle; accroître ces gains ne procure donc pratiquement pas plus d'utilité. Parallèlement, si les options rejetées ont des conséquences inacceptables, cela signifie que ces conséquences ont une désutilité marginale quasi infinie; les pertes d'utilité qui y sont associées sont tellement importantes qu'elles ont une grande influence sur le choix de l'individu maximisateur malgré la faible probabilité qu'elles peuvent avoir de se réaliser. Aux yeux d'Harsanyi, étant donné qu'elle prescrit finalement des politiques sociales également compatibles avec la maximisation de l'utilité dans la plupart des situations pertinentes, la règle du maximin peut ainsi, au mieux, être une règle approximativement valide pour des applications particulières pratiques d'une théorie morale plus générale. (Harsanyi 1975, p. 606) Ce n'est toutefois certainement pas de cette façon que Rawls conçoit ses deux principes de justice lorsqu'il parle du « *caractère fondamental du choix d'une conception de la justice* ». (Rawls, 1987, p. 187)

Pour bien expliquer la nature du désaccord entre Rawls et Harsanyi, je propose de reprendre une distinction entre trois types de règles faite par Dudley Knowles dans le chapitre de son livre *Political Philosophy* consacré à l'utilitarisme (Knowles, 2001, p. 23-35). Premièrement, il y a les règles idéales : ce sont celles qui, à l'image de l'impératif catégorique de Kant ou des dix commandements du christianisme, sont inviolables et doivent être observées en toutes circonstances, sans exception. Deuxièmement, il y a les règles approximatives ou selon l'expression anglaise, les « *rules of thumb* » : ce sont des maximes, des adages ou des principes qui ne garantissent pas les meilleures conséquences dans tous les cas mais qu'il est tout de même généralement avantageux de suivre. Par définition, de telles

règles sont faillibles et enjoignent ceux qui les appliquent à faire preuve d'un certain discernement lorsque des exceptions s'imposent et que les règles doivent être brisées. Enfin, un troisième type de règles fort particulier est regroupé sous la dénomination des règles constitutives : ces règles, à l'image des déplacements de pièces permis aux échecs, n'ont de sens qu'au sein du contexte dans lequel elles ont été pensées et mises en place. En philosophie politique, le concept de règle constitutive trouve une application fertile dans le domaine des institutions sociales. Les règles institutionnelles sont nécessaires au bon fonctionnement des institutions sociales et, même si elles peuvent être contestées et remises en question globalement, elles exigent une stricte obéissance tant que le contexte de leur application demeure inchangé. Autrement dit, il est permis de se demander si les règles institutionnelles mises en place sont bel et bien les plus efficaces du point de vue de l'utilité sociale et même d'en proposer de meilleures, mais personne n'est autorisé à enfreindre impunément les règles en vigueur pour autant. De toute évidence, Rawls considère ses principes de justice non pas comme des règles approximatives mais plutôt comme des règles constitutives d'une association politique et économique mutuellement avantageuse entre personnes libres, égales et rationnelles.

2.7 Conclusion

Finalement, il appert que les discussions techniques liées à l'application de la théorie du choix rationnel au contexte de la position originelle profitent davantage au développement et au raffinement de théories utilitaristes comme celle de John Harsanyi. Cela étant dit, la critique utilitariste faite à Rawls amène ce dernier à préciser et clarifier ses arguments à l'encontre de l'utilitarisme, qui, comme je tenterai de le démontrer dans le prochain chapitre, peuvent être avancés indépendamment de la construction de la position originelle. Avant de continuer, j'aimerais toutefois revenir sur une différence importante entre l'analyse d'Harsanyi et celle de Rawls. Il s'agit du fait que la théorie morale utilitariste d'Harsanyi est une théorie éthique générale ou, pour reprendre l'expression de Rawls, une doctrine morale compréhensive alors que la théorie de la justice comme équité se veut une conception politique de la justice qui n'est valide que dans le contexte d'une entreprise de

coopération sociale. Ainsi, la théorie d'Harsanyi définit le juste en fonction du bien, alors que la théorie de Rawls procède dans l'ordre inverse. De plus, il est important de souligner que le choix des partenaires dans la position originelle porte sur des principes régissant la répartition des biens sociaux premiers. Ce qui est l'objet de l'application de la règle du maximin dans la théorie de Rawls n'est donc pas l'utilité moyenne des personnes qu'ils pourraient devenir mais bel et bien un indice de ces biens sociaux premiers. Cela ne désamorce pas la critique faite par Harsanyi selon laquelle le principe du maximin ne saurait être un principe de choix rationnel pour les partenaires dans la position originelle. Cependant, cela implique qu'il est possible que les partenaires, s'ils basent leur choix sur le principe de la maximisation de l'utilité attendue, optent pour des principes qui concordent avec l'application du critère du maximin à un indice de biens sociaux premiers comme celui proposé par Rawls. Ce qu'il est impératif de bien saisir ici, c'est que la maximisation de l'utilité est un principe du choix rationnel, tandis que le maximin est un critère de justice distributive qui peut être justifié, dans un certain contexte, à l'aide de la théorie du choix rationnel. Cette nuance me sera très utile pour tenter de fournir, dans le prochain chapitre, une justification utilitariste des deux principes de justice de John Rawls. La justification que je proposerai relève d'une perspective utilitariste indirecte, en plusieurs points semblables à celle d'Harsanyi mais qui se concentre sur l'utilité sociale des règles constitutives des institutions.

CHAPITRE III

LA RELATION ENTRE LA THÉORIE DE RAWLS ET L'UTILITARISME

3.1 Introduction

Dans le premier chapitre, j'ai présenté ce que j'estimais être les traits les plus importants de la conception de la justice comme équité de John Rawls, en insistant tout particulièrement sur l'argumentation à partir de la position originelle. Dans le deuxième, je me suis intéressé à une théorie rivale, la théorie utilitariste de John Harsanyi. Dans ce troisième et dernier chapitre, je m'efforcerai de clarifier la relation complexe qu'entretient la théorie de Rawls avec la doctrine utilitariste. J'insisterai encore une fois sur l'argumentation rawlsienne à partir de la position originelle et en examinerai quelques critiques utilitaristes. Je souhaite réussir à démontrer que, même s'il est exagéré de dire que la théorie de Rawls n'est ni plus ni moins qu'un utilitarisme sophistiqué, certaines ressemblances existent entre les deux théories. De plus, je soutiendrai que certains parallèles peuvent être établis entre l'argumentation d'un utilitariste comme Harsanyi et celle de Rawls. Par ailleurs, je m'éloignerai de l'argumentation de la position originelle l'espace d'une section, dans laquelle je m'intéresserai à la question de la stabilité des institutions. J'espère alors montrer que l'utilitarisme et le contractualisme de Rawls peuvent et doivent cohabiter dans le cadre de l'éthique publique d'une démocratie libérale.

Dans la prochaine section, je présenterai les arguments de deux utilitaristes, David Lyons et Sydney Alexander, qui estiment que la théorie de Rawls n'est, en bout de ligne, pas si éloignée de l'utilitarisme que ne le prétend l'auteur de *Théorie de la Justice*. J'expliquerai par ailleurs pourquoi il serait toutefois inexact de qualifier la théorie de Rawls d'utilitariste. Par la suite, dans une autre section, je me baserai sur des écrits de Thomas Pogge pour

qualifier l'approche de Rawls de semi-conséquentialiste. Puis, je tenterai de montrer à quoi pourrait ressembler une justification utilitariste des deux principes de justice de Rawls. Dans la section suivante, je parlerai des deux niveaux de justification dans la théorie de Rawls; celui où on justifie le choix des principes de justice au sein de la position originelle et celui où est justifiée la construction de la position originelle elle-même. Suivant T. M. Scanlon, je me demanderai ensuite si le contractualisme de Rawls est réellement bien servi par son argumentation à partir de la position originelle, qui ramène le choix des principes de justice au choix rationnel d'un seul individu représentatif. Dans une autre section, je traiterai des motivations morales internes et des raisons morales externes. J'aborderai alors la question de la justification des principes sous l'angle de la stabilité des institutions en m'appuyant sur un texte de Catherine Audard, dans lequel elle nuance la critique de Rawls à l'égard de l'utilitarisme en tant qu'éthique publique. Finalement, je résumerai quelques affinités et points de désaccord entre la théorie de Rawls, reprenant des idées présentées par Samuel Scheffler.

3.2 Rawls, un utilitariste?

Dans un article intitulé *Rawls vs Utilitarianism* paru peu de temps après *A Theory of Justice*, David Lyons identifie deux types d'arguments mis de l'avant par Rawls en faveur de ses deux principes de justice et contre le principe d'utilité. (Lyons, 1972, p. 535-545) D'un côté, on retrouve les arguments liés à la compatibilité de l'application des principes de justice avec certains de nos jugements moraux bien considérés. Par exemple, l'utilitarisme n'exclut pas d'emblée que puisse être justifié un système d'esclavage. Un problème majeur avec l'utilitarisme selon Rawls, c'est qu'il n'établit pas formellement une fois pour toutes les droits et libertés de base égaux pour tous comme étant fondamentaux. Au contraire, le premier principe de justice de Rawls garantit un système de droits et libertés le plus étendu possible. Il est donc impossible de justifier un système d'esclavage à l'aide des principes de justice de Rawls. Cela dit, et c'est le point sur lequel insiste Lyons, si certaines possibilités théoriques comme la justification de l'esclavage sont bel et bien présentes au sein d'une conception utilitariste de la justice, cela ne veut pas dire pour autant qu'elles aient quelque

chance que ce soit de se réaliser effectivement dans un contexte réaliste. En fait, il s'avère que sous certaines conditions, les implications pratiques de l'utilitarisme et celles de la théorie de Rawls sont fort semblables, voire même identiques. En outre, étant donné l'importance des droits et libertés de base sur le bien-être des individus, il est difficile d'imaginer qu'un utilitariste puisse réellement parvenir à justifier un système d'esclavage à moins de faire des hypothèses fort controversées sur la nature humaine. On pourrait, par exemple, essayer de soutenir qu'il existe d'immenses différences entre les êtres humains dans leur capacité à ressentir l'utilité. On pourrait également tenter de défendre l'idée que c'est dans la nature de certains hommes d'être esclaves, que c'est la meilleure vie qu'ils peuvent mener étant donnée leur infériorité morale. Or, de telles hypothèses heurtent évidemment notre sens commun et vont à l'encontre des valeurs libérales. Ce sont les hypothèses particulières de ce type qui sont problématiques et non la théorie morale utilitariste en tant que telle. L'argument de Rawls selon lequel ses deux principes de justice s'accordent mieux que le principe d'utilité moyenne avec nos jugements moraux bien considérés paraît donc finalement moins convaincant qu'il ne pouvait sembler à première vue.

Le deuxième type d'arguments identifié par Lyons rassemble les arguments contractualistes présentés par Rawls dans la position originelle. L'idée principale guidant ces arguments est que le principe d'utilité moyenne représente un choix trop risqué pour les partenaires, qui devraient donc préférer les deux principes de Rawls puisqu'ils établissent définitivement la priorité des droits et libertés de base et assurent que les inégalités socio-économiques profitent, en bout de ligne, aux plus défavorisés de la société. Lyons nous rappelle toutefois que Rawls précise que l'application de ses deux principes de justice requiert un certain contexte favorable. En effet, la priorité absolue des droits et libertés est établie à l'aide d'une théorie idéale qui présuppose que tous les individus se conformeront aux règles établies. Ainsi, l'atteinte d'un certain seuil de développement économique et social est notamment requise pour que l'on puisse passer de la conception générale de la justice proposée par Rawls à sa conception spéciale, qui renferme les deux principes. Si le choix des partenaires dans la position originelle s'effectue entre les deux principes de Rawls et le principe d'utilité moyenne, il est donc normal de supposer qu'un contexte favorable est aussi présent pour évaluer le principe utilitariste. Or, il est plausible que les risques que Rawls

associe au choix du principe d'utilité moyenne, bien qu'ils reposent sur des possibilités logiques, n'aient aucune réelle possibilité de se réaliser dans un contexte favorable. Si tel est le cas, opter pour le principe d'utilité moyenne n'est pas nécessairement plus risqué que de choisir les deux principes de Rawls. À ce sujet, j'ai d'ailleurs déjà mentionné que si l'hypothèse (fort plausible) de l'utilité marginale décroissante se vérifie, l'utilitarisme peut aller jusqu'à prescrire des politiques encore plus égalitaristes que les principes de justice de Rawls. Par ailleurs, si le contexte favorable évoqué par Rawls n'est pas vérifié, sa propre conception générale de la justice autorise que les droits et libertés de base soient sacrifiés au nom d'avantages socio-économiques. La prétendue différence entre les deux théories n'est donc pas limpide. Lyons estime encore une fois que l'argumentation de Rawls en faveur de sa théorie de la justice et contre l'utilitarisme n'est pas si pénétrante qu'elle pouvait le sembler à prime abord.

Certains auteurs, notamment Sydney Alexander, sont allés jusqu'à prétendre que la théorie de Rawls n'est finalement ni plus ni moins qu'une forme sophistiquée d'utilitarisme indirect :

Tandis que Rawls présente sa théorie comme une alternative à l'utilitarisme, je soutiendrai qu'il met de l'avant une théorie utilitariste particulière et que sa querelle n'est pas avec l'utilitarisme en tant que tel, mais seulement avec d'autres variantes de l'utilitarisme. (Alexander, 1974, p. 598)

Alexander soutient que la théorie de Rawls est utilitariste car ses principes sont dérivés du choix de personnes rationnelles poursuivant leurs intérêts (derrière un voile d'ignorance). Ainsi, Rawls semble s'accorder avec les utilitaristes pour dire que les institutions sociales les plus justes sont celles qui entraîneront les meilleures conséquences sur le bien-être des individus. Dire que la théorie de Rawls est conséquentialiste peut toutefois porter à confusion. En effet, ne l'oublions pas, la conception de la justice comme équité se veut d'abord et avant tout déontologique. Étant donné l'opposition classique faite en philosophie morale entre les théories éthiques conséquentialistes et les théories éthiques déontologiques, l'affirmation selon laquelle la théorie de Rawls a une dimension conséquentialiste exige que des clarifications supplémentaires soient apportées. J'y reviendrai dans la prochaine section.

Certes, l'argumentation de Rawls à partir de la position originelle, en ramenant le choix de principes de justice au choix rationnel d'un seul individu représentatif qui se préoccupe exclusivement de ses futurs intérêts, donne une teinte résolument utilitariste à sa théorie de la justice. Cela dit, l'interprétation d'Alexander selon laquelle la théorie de Rawls est utilitariste paraît erronée aux yeux de Thomas Pogge. Selon ce dernier, la conception de la justice présentée dans *Théorie de la justice* se distingue de l'utilitarisme puisqu'elle écarte de manière définitive la possibilité de déroger aux principes de justice dans l'optique d'un gain net d'utilité. Les deux principes de justice sont fondamentaux; ils ne sont pas subordonnés au critère de la maximisation de l'utilité. Pogge admet que dans *Two Concept of Rules*, un article paru en 1955, « *Rawls envisage un système social dont les institutions doivent être justifiées sur des bases utilitaristes mais dont les participants doivent se soumettre à ces pratiques (optimales) indépendamment de considérations utilitaristes.* » (Pogge, 1989, p. 40) Le problème auquel est alors confronté Rawls est celui d'expliquer comment la maximisation de l'utilité peut jouer un rôle crucial dans l'évaluation du caractère juste d'une pratique mais être complètement ignorée dans les décisions des individus qui doivent s'y soumettre. Dans *Justice as Fairness*, paru en 1958, et par la suite dans *Théorie de la justice*, Rawls surmonte cette difficulté en élaborant sa conception de la justice non pas à partir des bases utilitaristes mais bel et bien à partir d'idées contractualistes. Selon Rawls, ce ne sont pas les principes qui maximisent l'utilité totale ou moyenne qui sont justes mais plutôt les principes qui résulteraient d'un accord hypothétique entre personnes libres, égales et rationnelles. Une lecture utilitariste de l'argumentation de Rawls à partir de la position originelle tend à obscurcir cette précision. La façon particulière dont Rawls construit la position originelle et utilise la théorie du choix rationnel pour justifier le choix de ses deux principes n'est pas étrangère à cette problématique. La relation complexe qui existe entre la théorie de Rawls et l'utilitarisme contribue par ailleurs largement à embrouiller certains des enjeux soulevés dans *Théorie de la justice*. Je traiterai de ces questions dans des sections à venir.

Une lecture utilitariste de l'argumentation de Rawls à partir de la position originelle suggère qu'une justification conséquentialiste des deux principes, en faisant usage de la théorie du choix rationnel en situation d'incertitude, puisse assez naturellement signifier que les deux principes seront choisis si et seulement s'ils répondent au critère de la maximisation

de l'utilité attendue. Afin de respecter l'interprétation conventionnelle de la théorie du choix rationnel, on doit soutenir que des agents rationnels placés dans la position originelle opteront pour les principes moraux qui maximisent leur espérance d'utilité. Dans *La justice comme équité, une reformulation de Théorie de la justice*, Rawls admet d'ailleurs ceci : « l'argumentation guidée par la règle du maximin est compatible avec l'idée que les êtres rationnels maximisent leur utilité attendue, mais seulement si celle est conçue comme n'ayant aucun contenu substantiel.¹ » (Rawls, 2004, p. 141) Il admet donc que « les partenaires font usage d'une fonction d'utilité (je la nomme volontiers ainsi) construite de façon à refléter les conceptions normatives idéales utilisées pour organiser la justice comme équité. » (Rawls, 2004, p. 152). Cela ne veut toutefois pas dire que Rawls considère sa théorie de la justice comme équité comme une théorie utilitariste. La fonction d'utilité utilisée par Rawls pour représenter les préférences des partenaires « est fondée sur les besoins et les nécessités des citoyens – leurs intérêts fondamentaux – conçus comme de telles personnes; elle n'est pas fondée sur les préférences et les intérêts réels des gens. » (Rawls, 2004, p. 152) Rawls fait effectivement usage du concept formel d'utilité comme mesure ordinale reflétant le classement de préférences individuelles, mais il ne se considère pas pour autant utilitariste puisqu'il ne s'appuie pas sur la satisfaction maximale des préférences individuelles comme critère moral fondamental.

Il est intéressant ici de noter que les « préférences et intérêts réels » dont parle Rawls correspondent plus ou moins à ce que Harsanyi définit comme les « préférences manifestes » d'un individu, à distinguer des « préférences réelles » qu'il exprimerait dans des conditions idéales et suite à une réflexion appropriée basée sur toutes les informations pertinentes. Rappelons également qu'Harsanyi exclut de sa théorie morale toute considération relative à des préférences antisociales. Or, Rawls critique la manière un peu rapide et arbitraire par laquelle Harsanyi sélectionne les préférences dont on doit tenir compte et celles qu'il faut ignorer dans le calcul utilitariste :

Il nous doit une explication des raisons pertinentes pour considérer certains plaisirs, ou satisfactions, comme nuls et non avenus. [...] Nous avons besoin de savoir d'où viennent ses

¹ Ici, Rawls fait référence à la distinction que j'ai déjà faite entre l'utilitarisme expérientialiste et l'utilitarisme préférentialiste.

restrictions sur l'admission dans les fonctions d'utilité et comment elles sont justifiées. Tant qu'on ne répond pas à ces questions dans un cadre de pensée correctement spécifié et identifiable comme utilitariste, nous ne pouvons pas savoir si Harsanyi a le droit de les imposer. (Rawls, 2004, p. 143).

Harsanyi s'écarte donc de la vision utilitariste traditionnelle selon laquelle la maximisation de l'utilité implique la satisfaction de toutes les préférences effectives, peu importe leur source. Rawls ne voit pas quelles raisons utilitaristes pourraient justifier cette prise de distance. On peut alors aller jusqu'à se demander si la théorie d'Harsanyi est véritablement utilitariste. Dans la théorie de la justice de Rawls, une lecture éclairée montre que ce sont des arguments contractualistes qui déterminent la forme que revêt la fonction d'utilité attendue des représentants des citoyens, ce qui fait que le fondement de la justification de la théorie de la justice comme équité est résolument contractualiste et non utilitariste.

3.3 Le semi-conséquentialisme de Rawls

Revenons maintenant sur l'aspect conséquentialiste de la théorie de Rawls. Dans une section de son livre *Realizing Rawls* (Pogge, 1989, p. 36-47), Thomas Pogge explique que la théorie déontologique de Rawls peut également être qualifiée de conséquentialiste, mais seulement dans une certaine mesure et en un sens bien précis. L'argumentation de Rawls à partir de la position originelle suggère que la justification des deux principes soit essentiellement conséquentialiste. Cependant, les principes de justice de Rawls ont un caractère déontologique et demandent à être respectés sans tenir compte directement des conséquences que cela entraîne. Pour cette raison, Pogge qualifie à juste titre le conséquentialisme affiché par Rawls d'indirect. (Pogge, 1989, p. 42) Par ailleurs, dans l'argumentation à partir de la position originelle, la justification du choix des deux principes repose indirectement sur certaines conséquences du choix des partenaires mais pas sur toutes les variations de bien-être individuel. Pogge estime donc que la conception de la justice de Rawls, sans être « pleinement conséquentialiste », est néanmoins « généralement conséquentialiste ». L'approche de Rawls est ainsi, selon l'expression utilisée par Pogge, « semi-conséquentialiste ».

Par cela, il entend que Rawls s'accorde avec les utilitaristes pour dire que « *les bénéfices et les charges qu'une organisation sociale établit explicitement n'ont pas plus de poids moral que ceux qu'elle entraîne de manière prévisible.* » (Pogge, 1989, p. 45) C'est là l'aspect conséquentialiste de la conception déontologique de la justice de John Rawls; on ne peut écarter des considérations conséquentialistes sous prétexte que certaines conséquences n'étaient pas souhaitées alors qu'elles étaient pourtant prévisibles. Cependant, Rawls considère que « *les bénéfices et les charges qu'une organisation sociale engendre ont toujours un poids plus important que ceux qu'elle laisse simplement survenir.* » (Pogge, 1989, p. 45) En effet, dans la théorie de Rawls, seules les conséquences de la mise en place de la structure de base de la société importent. Ainsi, il appert que selon Rawls, la justice sociale ne consiste pas à corriger les inégalités naturelles; il s'agit plutôt de garantir que les inégalités engendrées par l'interaction sociale via la structure de base sont équitables, c'est-à-dire à l'avantage de tous et chacun. Comme l'explique Rawls, « *les institutions sociales ne doivent pas tirer avantage des contingences comme les dons innés, ou la position sociale de départ, ou la bonne ou la mauvaise fortune au cours de la vie, sauf selon des voies qui bénéficient à tous, y compris aux plus défavorisés.* » (Rawls, 2004, p. 173) Autrement dit, la distribution des biens naturels (tels la santé, les talents et les habiletés) ne fait pas partie de la conception de la justice de John Rawls. Celle-ci concentre plutôt son attention sur la répartition des biens sociaux premiers (tels les droits et libertés, le revenu et la richesse). La justification des deux principes de justice que présente Rawls dans la position originelle n'est donc pas, contrairement à ce que serait une justification utilitariste, pleinement conséquentialiste mais plutôt semi-conséquentialiste. En effet, une justification purement utilitariste impliquerait que toutes les variations du bien-être des individus soient considérées et non pas seulement celles imputables à la structure de base via la répartition des biens sociaux premiers. Je propose maintenant de regarder à quoi pourrait ressembler une telle justification utilitariste de la conception rawlsienne de la justice.

3.4 La possibilité d'une justification utilitariste des principes de justice de Rawls

Les indications de Rawls sur les motivations des partenaires dans la position originelle nous poussent à croire qu'ils sont des utilitaristes de la règle. Cela ne revient toutefois pas à dire, comme le fait Harsanyi, qu'ils choisiront conséquemment le principe de la maximisation de l'utilité moyenne comme guide pour le fonctionnement des institutions sociales de base. La distinction entre la procédure de décision et le résultat de son application est ici primordiale : le raisonnement des partenaires (la procédure de décision) placés dans la position originelle devra répondre à des critères utilitaristes si ces agents sont rationnels mais les principes qu'ils choisiront (le résultat de l'application de la procédure) ne sont quant à eux pas nécessairement directement utilitaristes. C'est ce qui permet d'entrevoir la possibilité d'une justification utilitariste de principes qui ne sont pas explicitement utilitaristes eux-mêmes. Cette possibilité est particulièrement pertinente pour la théorie de la justice comme équité considérée comme une conception politique de la justice devant faire l'objet d'un consensus par recoupement au sein d'une société pluraliste. Rawls explique dans *La justice comme équité, une reformulation de Théorie de la justice* :

Nous serions ravis si les utilitaristes pouvaient trouver, à partir de leur propre point de vue, un moyen d'adopter les idées et les principes de la justice comme équité. La réalisation de cette possibilité signifierait qu'ils peuvent se joindre à un consensus par recoupement sur cette conception. (Rawls, 2004, p. 152)

Ainsi, alors que des passages de *Théorie de la justice* pouvaient laisser croire que le projet de Rawls consistait à démontrer que l'utilitarisme est une doctrine morale erronée, les écrits ultérieurs de Rawls démontrent le contraire. L'intention de Rawls n'est pas de critiquer l'utilitarisme comme doctrine compréhensive : « *En tant que conception politique, la justice comme équité évite dans la mesure du possible d'émettre ce genre de critique.* » (Rawls, 2004, p. 152)

De nombreuses raisons peuvent expliquer pourquoi des principes autres que le principe d'utilité seraient préférés par des utilitaristes. Rawls lui-même en présente quelques-unes : « *notre connaissance limitée des institutions sociales* » et « *des circonstances actuelles* », « *les limites à apporter à la complexité des règles légales et institutionnelles* »,

« *la simplicité nécessaire dans les directives de la raison publique* ». (Rawls, 1995, p. 212). Bref, en raison notamment des contingences de la vie réelle et de la complexité des faits sociaux, une règle stipulant que les institutions sociales doivent poursuivre consciemment et sans relâche l'objectif de la maximisation directe de l'utilité sociale pourrait s'avérer contre-productive. Il vaut alors mieux opter pour des règles plus simples à appliquer pour guider le fonctionnement des institutions afin que celles-ci puissent réellement permettre d'atteindre, indirectement mais plus efficacement, l'objectif ultime que constitue la maximisation de l'utilité sociale. Comme l'explique Rawls :

Ces raisons ainsi que d'autres peuvent conduire l'utilitariste à penser qu'une conception politique de la justice dont le contenu est libéral est une approximation applicable satisfaisante, peut-être même la meilleure, de ce que le principe d'utilité, tout compte fait, demanderait. (Rawls, 1995, p. 212)

Autrement dit, les principes de justice de Rawls pourraient bien être les règles approximatives les plus à même de maximiser l'utilité en bout de ligne.

Un autre type de justification utilitariste des principes de justice rawlsiens me paraît toutefois plus intéressante que celle que je viens d'évoquer. Pour y arriver, il faut considérer les deux principes de justice de Rawls non pas comme des règles approximatives mais plutôt comme des règles constitutives desquelles on ne peut déroger sous aucun prétexte. Plus précisément, ce sont des règles constitutives d'une association entre individus libres, égaux et rationnels qui ont déjà déterminé qu'il était dans leur intérêt de s'associer. On se rapproche ainsi de l'idée fondamentale qui inspire la théorie de Rawls. Toutefois, plutôt que de considérer les règles justes comme étant celles sur lesquelles s'entendraient des personnes rationnelles placées derrière un épais voile d'ignorance et ne pouvant procéder à des raisonnements probabilistes, la perspective utilitariste nous amène alors à considérer, comme Harsanyi, le point de vue de l'observateur impartial qui raisonne comme s'il avait une probabilité égale d'être chaque personne. L'observateur impartial cherche donc à satisfaire autant que possible les préférences individuelles en maximisant l'utilité moyenne au sein de la société. Toutefois, il cherche à le faire indirectement, via la mise en place de règles contraignantes. On peut ici rejoindre Rawls pour dire que ces règles constitutives se limitent aux principes qui régissent le fonctionnement des institutions sociales les plus importantes (la

constitution politique et les principales structures socio-économiques). Ainsi, le choix de l'observateur impartial porte en quelque sorte sur les « règles du jeu » qui encadrent la coopération sociale. La question qui lui est posée est la suivante : quelles sont les meilleures règles ou autrement dit les règles qui permettent d'atteindre le plus haut niveau d'utilité sociale dans le cadre d'un jeu coopératif étant donné ce que l'on sait de la psychologie humaine et des conditions sociales actuelles? Une autre façon de poser la question revient à ceci : quelles règles constitutives de leur association cadrent le mieux avec les préférences morales d'un individu représentatif d'une société libérale contemporaine?

La réponse ne va pas de soi. Il s'agit en fait d'une question fort complexe à laquelle plusieurs réponses pertinentes pourraient être apportées. La variété des réponses possibles découle de la pluralité des valeurs à prendre en considération et des innombrables combinaisons que l'on peut obtenir en leur accordant des poids relatifs variables. Ainsi, certains insisteront davantage sur la valeur de l'égalité, d'autres sur celles de la liberté ou du respect de l'individu, d'autres enfin sur l'efficacité collective. Il me semble par ailleurs que la combinaison des deux principes de justice de Rawls représente une réponse intéressante. Ce n'est pas tellement surprenant puisque Rawls admet volontiers que les valeurs principales qui animent sa conception de la justice sont largement partagées au sein des démocraties libérales contemporaines. Ainsi, même si Rawls lui-même n'apprécierait sûrement pas une telle formulation qui implique une forme de subjectivisme, les préférences morales dont ses deux principes de justice rendent compte sont sans doute représentatives de celles d'une bonne part de la population de nos sociétés libérales.

Étant donné l'importance que les droits et libertés de base ont pour le bien-être des individus, un principe proche du premier principe rawlsien, celui de la plus grande liberté égale pour tous, devrait être justifié par toute théorie utilitariste sérieuse. L'immense perte d'utilité associée au non-respect des droits et libertés fondamentaux des individus paraît en effet trop importante pour que ceux-ci ne soient pas protégés avec soin. En fait, il est difficile, voire même impossible, d'imaginer comment, dans un contexte réaliste, un gain net d'utilité totale pourrait être réalisé en enfreignant les droits et libertés de base des individus. Pour éviter toute ambiguïté dans l'esprit des citoyens et décourager toute possibilité d'abus, il peut

alors être utile d'établir formellement la priorité des ces droits et libertés de base. Je crois que la première partie du second principe, qui porte sur la juste égalité des chances, peut également être justifié assez aisément, quoi que moins fermement, sur la base du principe d'utilité. En effet, le fait que les diverses positions sociales soient ouvertes à tous est fort probablement le meilleur moyen de s'assurer que les différents postes seront comblés par les individus les plus compétents pour réaliser les tâches qui y sont associées. Pour qu'une dérogation au principe de la juste égalité des chances soit justifiée d'un point de vue utilitariste, il faudrait donc que cela implique des avantages considérables liés à des considérations d'efficacité. Ce n'est pas exclus d'emblée, mais si l'on tient également compte du fait que les possibilités de carrière et les perspectives de vie qu'envisagent des individus libres sont elles-mêmes une source importante d'utilité, il serait surprenant que la maximisation de l'utilité exige que l'on s'écarte d'une juste égalité des chances. Il peut alors être préférable, d'un point de vue utilitariste, d'établir une règle assurant le respect d'une juste égalité des chances prioritairement à l'application du calcul utilitariste à des cas précis. Le seul problème, et il est de taille, est alors d'établir les limites de ce qui constitue une « juste » égalité des chances. Un peu de la même manière, il faudrait également définir plus clairement quelles sont les droits et libertés de base devant faire partie du premier principe. Par ailleurs, on peut considérer qu'une juste égalité des chances est un droit de base, ce qui permettrait d'incorporer le principe de la juste égalité des chances au premier principe. Je n'ai malheureusement ni le temps ni l'espace pour élaborer sur ces épineuses questions qui sont par ailleurs tout aussi pertinentes pour la théorie de Rawls.

Pour ce qui est du principe de différence, une justification utilitariste me semble moins facile à trouver. Un principe qui garantirait un revenu minimum couvrant les besoins fondamentaux me semble plus facilement justifiable d'un point de vue utilitariste. En effet, on n'a qu'à soutenir que ce minimum social, indispensable pour avoir la chance d'être un véritable membre coopérant de la société, apporte à chaque personne démunie une utilité presque infiniment plus élevée que la différence que ce montant supplémentaire pourrait apporter à qui que ce soit d'autre. Encore une fois, on pourrait également inclure le droit à un revenu minimum dans les droits de base. Dans *Libéralisme politique*, Rawls admet d'ailleurs ceci : « *quoiqu'un minimum social couvrant les besoins fondamentaux de tous les citoyens*

fasse [...] partie des questions constitutionnelles essentielles, ce que j'ai appelé le « principe de différence » est plus exigeant et n'en fait pas partie » (Rawls, 1995, p. 278) Comme l'explique Rawls dans la préface de l'édition française de Théorie de la justice, l'argumentation « en faveur du principe de différence implique un équilibre de considérations plus délicat. » (Rawls, 1987, p. 12)

Il convient par ailleurs de préciser que Rawls, en disant cela, ne considérait aucunement qu'une justification utilitariste de ses deux principes soit plus pertinente que la justification contractualiste qu'il propose. L'avantage d'une telle justification utilitariste vient de ce qu'elle ramène tous nos jugements moraux à un seul et unique critère normatif, celui de la maximisation de l'utilité. Or, pour une conception politique de la justice devant faire l'objet d'un consensus par recoupement dans une société pluraliste, cet avantage n'en est pas un, au contraire. Une justification contractualiste est mieux appropriée à ce contexte. Avant d'aller plus loin, il convient de traiter brièvement des deux niveaux de justification que l'on retrouve au sein de la théorie de la justice de Rawls afin de clarifier certaines idées qui demeurent pour l'instant embrouillées, ce que je ferai dans la prochaine section.

3.5 Les deux niveaux de justification dans la théorie de la justice de Rawls

Il existe, dans la théorie de la justice comme équité de John Rawls, deux niveaux de justification que l'on peut distinguer mais qui demeurent intimement liés. Dans un premier temps, il y a la justification du choix des deux principes au sein de la construction de la position originelle. Elle consiste à expliquer pourquoi, dans le contexte de la position originelle, les partenaires sont amenés à s'entendre sur les principes de justice de Rawls pour régir le fonctionnement des principales institutions sociales. Dans un deuxième temps, il y a la justification de la construction de la position originelle elle-même ou, autrement dit, des conditions particulières dans lesquelles le choix des principes de justice doit s'effectuer. Pour ce qui est du premier niveau, rappelons brièvement les trois principaux arguments présentés par Rawls dans son argumentation à partir de la position originelle en faveur du maximin et contre le principe de la maximisation de l'utilité moyenne. Premièrement, nous dit Rawls, les

partenaires étant placés derrière un voile d'ignorance, n'ont aucune base objective pour établir des raisonnements probabilistes. Quant au principe de raison insuffisante, qui permettrait d'attribuer une chance égale de se retrouver dans chacune des positions sociales envisageables, son application serait inappropriée dans la position originelle, selon Rawls. L'importance de ce qui est en jeu et le caractère risqué que comporterait l'application du principe de raison insuffisante dans le contexte particulier de la position originelle explique pourquoi les partenaires devraient l'exclure de leur raisonnement. Le deuxième argument évoqué par Rawls est lié au fait que ses deux principes de justice assurent un minimum satisfaisant aux partenaires, qui n'ont peu ou pas d'intérêt pour des gains allant au-delà de ce minimum satisfaisant. Enfin, Rawls soutient que le principe d'utilité moyenne pourrait avoir des conséquences intolérables que certaines personnes ne pourraient accepter.

C'est sur ce type de justification de premier niveau à laquelle est consacrée l'argumentation à partir de la position originelle que j'ai concentré la majeure partie de mon attention jusqu'à maintenant. Comme j'ai tenté de le montrer, les arguments de Rawls en faveur de l'application de la règle de décision du maximin, qui semblent notamment justifier le choix du principe de différence, posent quelques problèmes que Rawls a du mal à résoudre à partir de la position originelle sans créer de tension au sein de sa théorie. Cela dit, Rawls apporte une précision importante à ce sujet dans *La justice comme équité, une reformulation de Théorie de la Justice* :

En dépit de la ressemblance formelle entre le principe de différence et la règle du maximin [...], le raisonnement en faveur du principe de différence ne fait pas appel à cette règle. La ressemblance formelle est trompeuse. [...] Le défaut d'explication de ce point constitue un grave défaut de TJ. (Rawls, 2004, p. 136)

Plutôt que de soutenir que les partenaires dans la position originelle fondent leur choix sur la règle du maximin, il convient donc de dire qu'ils font un choix qui est compatible avec l'application de cette règle ou qui peut être interprété de cette façon : « *il n'est pas essentiel pour les partenaires d'utiliser la règle du maximin dans la position originelle. Il s'agit simplement d'un dispositif heuristique commode.* » (Rawls, 2004, p. 142)

Dans *La justice comme équité, une reformulation de Théorie de la justice*, Rawls affirme en outre que les orientations données par la règle du maximin fournissent la base d'une argumentation « *assez décisive pour appuyer les droits et libertés de base égaux. En revanche, ces orientations apportent peu de soutien au principe de différence.* » (Rawls, 2004, p. 138) De plus, Rawls semble reconnaître les difficultés relatives au premier argument, qui exclut le recours à tout raisonnement probabiliste. En effet, pour défendre le choix du premier principe au détriment du principe d'utilité moyenne, il préfère désormais insister sur les deux autres arguments, ceux du minimum satisfaisant associé au premier principe et des conséquences intolérables que pourrait entraîner l'application du principe utilitariste. Ces deux arguments combinés font en sorte qu'il est rationnel pour les partenaires d'établir définitivement une liste de droits et libertés inviolables plutôt que de soumettre constamment ces derniers au calcul utilitariste. Bien entendu, comme je l'ai déjà souligné, il est difficile d'imaginer comment le principe d'utilité moyenne pourrait justifier le non-respect des droits et libertés de base dans des conditions réalistes favorables. Mais pourquoi prendre le moindre risque si cela ne peut amener rien de bon mais a le potentiel d'entraîner des conséquences intolérables? Compte tenu du fait du pluralisme raisonnable, il s'avère particulièrement important, dans une société démocratique contemporaine, de « *fixer une fois pour toutes les droits et libertés de base et de leur assigner une priorité spéciale.* » (Rawls, 2004, p. 161) De plus, les implications du premier principe de justice de Rawls sont évidemment plus claires et précises que celles du principe d'utilité moyenne, ce qui lui confère un avantage en tant que raison publique de justification. (Rawls, 2004, p. 162) Enfin, Rawls estime que le fait d'établir la priorité absolue des droits et libertés de base, en éliminant du débat politique des questions hautement conflictuelles, est plus à même d'encourager le développement des « *vertus coopératives de la vie politique.* » (Rawls, 2004, p. 163)

Pour leur part, les raisons spécifiques en faveur du choix du principe de différence relèvent des idées de publicité, de réciprocité et de stabilité. Ces idées sont liées au concept de raison publique au sein d'une société bien ordonnée. Rawls considère que la condition de publicité, en exigeant que les raisons mises de l'avant dans le débat public puissent être reconnues par toute personne raisonnable sur une base commune, « *confère un rôle éducatif à*

la conception politique de la justice. » (Rawls, 2004, p. 170) Si ce rôle est bien rempli par la conception politique de la justice comme équité, les citoyens se concevront comme des êtres libres et égaux engagés dans une entreprise de coopération mutuellement avantageuse. Ils pourront alors apprécier comment l'idée plus profonde de réciprocité est incorporée dans le principe de différence puisque ce dernier garantit que « *les contingences* (sociales, naturelles et fortuites) *sont exploitées seulement de manière à profiter à tout le monde.* » (Rawls, 2004, p. 175) Finalement, ceci aura l'effet bénéfique, au niveau de la stabilité du système de coopération, d'encourager encore davantage « *la confiance mutuelle et les vertus coopératives* » (Rawls, 2004, p. 175) Par ailleurs, le principe de différence garantit un minimum social basé sur une idée de réciprocité qui va au-delà de ce qu'exigerait un principe d'utilité restreinte garantissant à chaque citoyen un minimum couvrant ses besoins essentiels préalablement à la maximisation de l'utilité moyenne. Le minimum social envisagé par Rawls « *couvre au moins les besoins essentiels à une vie décente, et probablement plus.* » (Rawls, 2004, p. 180) En tenant compte des liens de l'engagement, le principe d'utilité, en assurant un minimum social plus satisfaisant que le principe d'utilité restreinte, est ainsi plus susceptible de maintenir la stabilité du système coopératif.

Comme nous l'avons vu, l'argumentation à partir de la position originelle fait usage de la théorie du choix rationnel en situation d'incertitude pour expliquer le raisonnement des partenaires. Ainsi, le premier niveau de justification dans la théorie de Rawls repose en bonne partie sur une application particulière de la théorie de la décision rationnelle pour expliquer le choix des principes de justice par les partenaires. Cela dit, les arguments de Rawls pour justifier le choix des partenaires dans la position originelle ne sont pas inhérents à la théorie de la décision rationnelle, mais sont plutôt reliés à d'autres parties de *Théorie de la justice*. Ainsi, on ne peut saisir tout le poids et la portée de ces arguments qu'en allant au-delà de l'argumentation formelle à partir de la position originelle et en considérant de plus près les questions soulevées par Rawls au sujet de la stabilité d'un système social, mais également celles du respect de soi et des liens de l'engagement. C'est pourquoi la justification du choix des partenaires dans la position originelle ne saurait être complètement indépendante de la justification de la position originelle elle-même.

À un deuxième niveau, on retrouve donc la justification de la position originelle elle-même, en tant que situation hypothétique appropriée pour juger du caractère moral des principes de justice sociale. La position originelle est un procédé de représentation qui remplit deux rôles principaux dans la théorie de la justice de John Rawls. Le premier est de modéliser les « *conditions équitables dans lesquelles les représentants des citoyens, conçus uniquement comme des personnes libres et égales, doivent s'accorder sur les termes équitables de la coopération sociale.* » (Rawls, 2004, p. 117) Le second est de modéliser « *les limitations équitables des raisons sur la base desquelles les partenaires (en tant que représentants des citoyens) situés dans ces conditions équitables peuvent légitimement accepter certains principes de justice et en rejeter d'autres.* » (Rawls, 2004, p. 117-118) Par ailleurs, en permettant de garder la trace des hypothèses faites sur la description des partenaires et de la situation dans laquelle ils se retrouvent, la position originelle a l'avantage d'unifier ces hypothèses « *au sein d'une idée facile à saisir qui nous permet d'envisager leurs implications plus aisément.* » (Rawls, 2004, p. 118) Essentiellement, le deuxième niveau de justification garantit que la manière dont sont sélectionnés les principes de justice est elle-même juste au sens de la théorie de la justice comme équité. En effet, la construction de la position originelle encadre le processus décisionnel des partenaires pour s'assurer que celui-ci repose sur des règles de décision qui susciteraient elles-mêmes l'accord unanime de personnes libres et rationnelles souhaitant s'entendre sur les termes de leur coopération sociale. Le raisonnement des partenaires doit ainsi se soumettre à des principes jugés raisonnables lorsqu'il s'agit de délibérer sur la justice sociale. C'est ainsi que Rawls rejette la règle de la maximisation de l'utilité ainsi que la logique sacrificielle qui lui est implicite, soutenant qu'une telle logique est inadéquate pour déterminer comment répartir équitablement les biens sociaux premiers.

Jusqu'ici, je n'ai qu'effleuré les questions relatives à ce deuxième niveau de justification, ce qui ne veut pas dire qu'elles sont sans importance pour autant. En fait, je crois que la justification profonde de la théorie de la justice comme équité se retrouve à ce deuxième niveau. J'ai d'ailleurs souligné plus tôt que la construction de la position originelle était cruciale dans la justification complète des deux principes car ce sont ses caractéristiques qui garantissent que l'accord entre les partenaires sera équitable. Ainsi, alors qu'une

justification utilitariste du choix des principes de justice rawlsiens dans le contexte de la position originelle est envisageable, la justification de la construction de la position originelle est résolument contractualiste dans la théorie de Rawls. Dans une perspective purement utilitariste, la position originelle ne saurait être rien de plus qu'une fiction utile permettant aux individus de trouver une motivation d'agir en fonction de règles dont ils seraient incapables d'apprécier la pertinence autrement. Dans la perspective de Rawls, nous avons vu que la position originelle est plus qu'une fiction utile puisqu'elle joue un rôle important dans l'entreprise de justification de la théorie de la justice comme équité.

Par ailleurs, je n'ai pour l'instant pas véritablement remis en question la présupposition suivante : des principes justes sont ceux qu'un être rationnel ignorant la position qu'il occupe dans la société choisirait. Un problème majeur de l'argumentation de Rawls à mes yeux est lié au fait que la position originelle est construite de telle sorte que le contenu d'un accord entre une multitude de personnes aux intérêts divergents est ramené au choix individuel d'un seul agent représentatif. La question qui se pose maintenant est de savoir si le projet contractualiste de Rawls est réellement bien servi par la construction de sa position originelle. Autrement dit, l'idée fondamentale de la théorie de Rawls, celle d'un accord entre personnes libres, égales et rationnelles sur des principes de justice, est-elle bien illustrée par un processus de choix individuel, même si celui-ci est impartial? C'est l'objet de la prochaine section.

3.6 Le contractualisme de Rawls et la position originelle

Malgré le fondement contractualiste explicite que donne Rawls à sa théorie de la justice comme équité, le choix des principes de justice dans la position originelle est ramené au choix rationnel d'un seul individu représentatif. Pourtant, un reproche que Rawls adresse à l'utilitarisme est justement « *d'adopter pour la société dans son ensemble le principe du choix rationnel valable pour un individu* ». (Rawls, 1987, p. 52) En agrégeant toutes les préférences individuelles dans celles d'un spectateur impartial imaginaire, l'utilitarisme commet selon Rawls l'erreur de tenter de fondre des systèmes de désirs conflictuels en un

système unique et cohérent. Ainsi, l'utilitarisme présente les individus comme de simples producteurs de satisfaction séparés et commande ensuite une gestion efficace des ressources afin de maximiser la somme des satisfactions individuelles agrégées. La fiction du spectateur impartial et sympathique permet alors de traiter toutes les personnes comme une seule, ce qui amène Rawls à affirmer : « *La pluralité des personnes n'est donc pas vraiment pris au sérieux par l'utilitarisme.* » (Rawls, 1987, p. 52)

Pour s'assurer que le choix de l'agent représentatif dans la position originelle n'aboutira pas à un principe approprié pour un seul individu rationnel, comme le principe de la maximisation de l'utilité moyenne des personnes qu'il pourrait devenir, mais plutôt à des principes auxquels aucun être rationnel ne pourrait raisonnablement refuser d'adhérer, Rawls doit exclure d'emblée les raisonnements probabilistes qui ouvriraient la voie à une logique agrégative. J'estime qu'il affaiblit du même coup le premier niveau de justification qui dit s'appuyer sur la théorie de la décision rationnelle afin de demeurer fidèle au choix qu'il a fait au deuxième niveau de justification, c'est-à-dire au moment de la construction de la position originelle. N'aurait-il donc pas été préférable que Rawls décrive explicitement ses principes de justice comme étant ceux qu'aucun être rationnel ne pourrait refuser, comme le suggère la théorie plus générale de T. M. Scanlon, lui donnant ainsi l'allure du résultat d'une négociation rationnelle informée plutôt que celle d'un choix rationnel individuel en situation d'incertitude? En ramenant le contenu d'un accord entre personnes ayant des intérêts divergents au choix d'un seul individu, Rawls s'écarte du fondement contractualiste de sa théorie de la justice. En s'appuyant sur la théorie du choix rationnel pour défendre ses deux principes de justice à partir de la position originelle et rejeter le principe utilitariste par la même occasion, et ce tout en plaçant les partenaires dans une situation d'incertitude mais en refusant de s'appuyer sur la théorie bayésienne moderne de la décision rationnelle, Rawls crée à tout le moins une tension certaine au sein de sa propre théorie de la justice.

En effet, Rawls présente l'hypothèse de rationalité des partenaires de manière tout à fait conventionnelle lorsqu'il affirme : « *L'hypothèse que les partenaires sont rationnels et mutuellement désintéressés revient alors à ceci : les personnes placées dans la position originelle essaient de reconnaître des principes qui favorisent autant que possible leurs*

systèmes de fins. » (Rawls, 1987, p.176) Il s'agit donc, comme nous l'avons déjà souligné, d'une rationalité de type instrumentale, presque identique à celle dont les agents sont dotés dans les modèles élaborés par les économistes. Un individu rationnel, nous dit Rawls, classe les différentes options qui s'offrent à lui en fonction des buts qu'il s'est fixés et entreprend les actions qui satisfont le mieux ses préférences (Rawls, 1987, p.175). Voilà le genre de propos qui aurait très bien pu sortir de la bouche d'un utilitariste convaincu! En se basant sur la théorie du choix rationnel pour rejeter l'utilitarisme, Rawls s'aventure sur un terrain miné. En effet, comme le souligne Brian Barry², même si Rawls a du mal à l'admettre, la logique de la construction de la position originelle pointe naturellement vers la conclusion que l'individu représentatif de l'ensemble des partenaires cherche d'une certaine manière à maximiser quelque chose. Or, ce quelque chose saurait difficilement être autre que l'utilité moyenne des personnes qu'il peut devenir (i.e. l'utilité attendue moyenne). Les restrictions supplémentaires que Rawls doit faire pour s'assurer que le choix de l'individu représentatif aboutira aux deux principes tendent à dénaturer le problème de choix interprété dans les termes conventionnels de la théorie du choix rationnel. Il en résulte que la façon dont Rawls tente d'écarter le principe d'utilité en faisant appel à des prémisses qui peuvent être qualifiées d'essentiellement utilitaristes dévoile une certaine ambiguïté ou tension interne qui contribue à obscurcir plutôt qu'à clarifier les arguments en faveur de ses deux principes de justice.

Dans *Contractualism and Utilitarianism* (Scanlon, 1982), T.M. Scanlon établit une distinction fort pertinente entre ce qu'il appelle l'utilitarisme philosophique et l'utilitarisme normatif. L'utilitarisme philosophique, nous explique Scanlon, est « *une thèse philosophique particulière à propos du sujet d'intérêt de la moralité, à savoir la thèse que les seuls faits moraux fondamentaux sont des faits à propos du bien-être individuel.* » (Scanlon, 1982, p.108) L'utilitarisme normatif regroupe quant à lui toutes les doctrines prescrivant, comme principe normatif, la maximisation de l'utilité (totale ou moyenne, directement ou indirectement). De manière assez naturelle, l'utilitarisme philosophique conduit à une forme ou une autre d'utilitarisme normatif (utilitarisme classique, utilitarisme moyen, utilitarisme

² « *The form of the decision problem facing the people in the original position is naturally construed as one of maximising under conditions of uncertainty. And this leads us (as with Harsanyi) to the conclusion that the principle chosen should be one of maximizing the average utility of the people they might turn out to be.* » (Barry, 1989, p. 339)

de l'acte, utilitarisme de la règle, etc.). Il est par ailleurs possible d'aboutir à des principes normatifs utilitaristes par une autre voie. Ainsi, comme le fait Harsanyi, il est possible de dériver le principe d'utilité moyenne, à partir d'une construction apparemment contractualiste, comme le résultat du choix de partenaires devant s'entendre sur les termes de leur association. Cela dit, Scanlon estime que le succès et l'influence des principes normatifs utilitaristes s'expliquent en bonne partie par l'attrait indéniable de l'utilitarisme philosophique. (Scanlon, 1982, p. 108) Or, aux yeux de Scanlon, le contractualisme propose une thèse rivale supérieure à l'utilitarisme philosophique quant à la nature de la moralité et c'est sur ce point qu'il faut insister pour déclasser l'utilitarisme sous toutes ses formes. Selon Scanlon, les revendications reliés au bien-être individuel représentent une classe d'arguments moraux valides, mais elles ne sauraient être la seule. (Scanlon, 1982, p. 108) De plus, sa propre version du contractualisme, qui définit la justice en termes de principes « *que personne ne pourrait raisonnablement rejeter* » (Scanlon, 1982, p. 111), n'aboutit pas à des principes normatifs utilitaristes.

Scanlon remet par ailleurs en question l'argument de Harsanyi en faveur du principe d'utilité moyenne et, par la même occasion, il critique la manière dont Rawls construit sa position originelle. Pour ce faire, Scanlon propose de découper l'argument de Harsanyi en trois étapes. (Scanlon, 1982, p. 121-124) La première étape consiste à affirmer que des principes, pour être moraux, doivent être acceptables d'un point de vue impartial. La deuxième étape est de dire que le point de vue impartial par excellence est celui d'une personne qui choisit en ignorant quelle position elle occupe dans la société. Enfin, la troisième étape renferme l'idée d'un choix rationnel effectué en considérant une probabilité égale d'occuper n'importe quelle position. Il y a donc deux pas à franchir dans l'argument d'Harsanyi : celui permettant de passer de la première étape à la deuxième et celui menant de la deuxième à la troisième. Rawls fait le premier pas avec Harsanyi puisqu'il accepte l'idée que ses principes de justice soient le résultat du choix d'un individu rationnel placé derrière un voile d'ignorance. Rawls refuse toutefois de faire le second pas car il écarte le recours à toute forme de raisonnement probabiliste. L'erreur de Rawls, selon Scanlon, est d'avoir accepté de faire le premier pas avec Harsanyi; dès lors qu'il ramène le contenu de l'accord originel à un choix rationnel individuel, Rawls s'écarte de la perspective contractualiste qui

constitue pourtant l'idée fondamentale derrière sa théorie de la justice et se rapproche du point de vue utilitariste de l'observateur impartial. La tension découle d'interprétations divergentes de ce en quoi consistent exactement la justice, l'impartialité et l'équité. Les deux niveaux de justification dans la théorie de Rawls entrent alors en conflit puisque au premier niveau, on tente de justifier le choix de principes de justice par la décision rationnelle d'un seul agent motivé exclusivement par ses intérêts tandis qu'au deuxième niveau, on essaie de s'assurer que les principes choisis seront tels qu'aucune personne rationnelle ne pourrait raisonnablement refuser d'y adhérer. Qu'est-ce qui motive réellement le choix des partenaires alors? Leurs intérêts personnels, comme le veut le premier niveau de justification? Ou le désir de s'entendre sur les termes d'un accord raisonnable et équitable, comme le stipule le deuxième niveau?

D'après Scanlon, les principes que personne ne pourrait raisonnablement rejeter ne sont pas nécessairement les mêmes que les principes qu'une personne rationnelle choisirait sans savoir quelle position elle occupe dans la société; ce sont là deux questions fort différentes. Bien que Rawls propose à quelques reprises la première formulation dans *Théorie de la Justice* (celle d'un accord auquel aucun être rationnel ne pourrait raisonnablement refuser d'adhérer), il privilégie finalement la seconde option (celle d'un seul individu représentatif choisissant derrière un voile d'ignorance) lorsque vient le temps de décrire le raisonnement des partenaires dans la position originelle. Suivant Scanlon, j'estime que ce choix est malheureux. Cela dit, Rawls demeure tout de même, dans l'ensemble, fidèle à une justification de type contractualiste puisqu'il incorpore les questions relatives à la stabilité du système de coopération et aux liens de l'engagement dans le raisonnement des partenaires. Il insiste par le fait même sur le caractère spécial du choix des partenaires, qui porte sur les termes d'un accord fondamental. Ainsi, même s'ils sont au départ motivés uniquement par leurs intérêts, les partenaires insistent également pour s'assurer que les principes choisis seront acceptables pour tous. Seulement, il me semble que cette manière détournée de procéder ne contribue pas avantageusement à la clarté de la justification contractualiste de la théorie de Rawls. À mon avis, il aurait été plus simple et plus clair de dire que les partenaires sont motivés d'abord et avant tout par le désir de s'entendre sur les termes d'un accord équitable et non pas exclusivement par leurs intérêts privés.

Par contre, si on admet que le rôle de la position originelle est essentiellement heuristique dans la théorie de Rawls, c'est à l'ensemble des lecteurs de Rawls de juger de la pertinence de la construction théorique qu'elle représente. Or, il se peut que les problèmes que j'ai soulignés avec l'argumentation de Rawls ne soient pas d'une grande importance pour des gens qui connaissent peu la théorie du choix rationnel en situation d'incertitude ou qui ne s'intéressent pas aux plus récents développements de l'utilitarisme. Par ailleurs, contrairement à Scanlon et à ce qu'il avait pu lui-même prétendre dans *Théorie de la justice*, rappelons que le projet global de Rawls n'est pas de démontrer la supériorité du conséquentialisme par rapport à l'utilitarisme pour aborder l'ensemble des questions morales. Ce n'est finalement qu'en tant que conception politique de la justice jouant le rôle d'éthique publique d'une démocratie libérale que Rawls condamne l'utilitarisme. Je me pencherai maintenant sur les critiques spécifiques que Rawls adresse à l'utilitarisme à ce sujet.

3.7 Motivations internes et raisons externes

Une critique importante que fait Rawls à l'égard de l'utilitarisme concerne le type de justification requis pour les institutions d'une démocratie libérale contemporaine. En effet, selon Rawls, l'utilitarisme ne saurait être accepté en tant qu'éthique publique d'une société démocratique pluraliste. Ainsi, comme l'explique Catherine Audard, Rawls « *critique le type de « raisons » que l'utilitarisme favorise* » (Audard, 2002, p. 50) L'utilitarisme, concentrant son attention sur la délibération personnelle des individus, s'intéresse essentiellement aux motivations individuelles ou, autrement dit, aux raisons internes. Or, selon Rawls, dans le débat public, seules les raisons publiques doivent être admises : « *une bonne raison, si elle n'est pas publique, n'est pas une raison valide.* » (Audard, 2002, p. 59) Contrairement aux raisons internes de l'utilitarisme, les raisons publiques de Rawls sont indépendantes des motivations personnelles déjà existantes. Cette caractéristique constitue selon lui un point faible de l'utilitarisme, qui fait l'erreur de privilégier les préférences de la majorité, transformant le système politique ni plus ni moins qu'en une « démocratie de marché »³.

³ L'expression est, originalement, de Ronald Dworkin.

On peut toutefois se demander si, dans sa conception de l'éthique publique, Rawls n'a pas tort d'exclure systématiquement les raisons internes de l'utilitarisme. Sur cette question, Catherine Audard se positionne contre Rawls :

Pour Rawls, présenter des raisons internes va contre la conception de l'impartialité qu'il suppose et qui est construite sur une distinction entre la sphère privée et la sphère publique où seule on peut parler de raisons (il n'y a pas de raisons privées, dit-il). Contre Rawls, je voudrais conclure en disant qu'une éthique publique dans une démocratie doit combiner raisons publiques et raisons internes, tout en reconnaissant la tension qui existe entre elles. (Audard, 2002, p. 60)

Le défi posé par la question de la stabilité des institutions n'est-il pas justement de trouver des principes de justice que des êtres rationnels (motivés par les raisons internes que constituent leurs intérêts) et raisonnables (sensibles aux raisons externes fournies par des considérations de justice et d'équité) pourraient endosser et continuer à accepter sur une longue période de temps? Si tel est le cas, les approches contractualistes et utilitaristes, malgré toutes les difficultés que cela comporte, doivent non pas se confronter mais se compléter au sein de l'éthique publique d'une société démocratique libérale. Catherine Audard est de cet avis :

Les démocraties contemporaines doivent donc être à la fois utilitaristes et contractualistes, fondées à la fois sur des raisons publiques et sur des raisons internes. C'est là leur force, comme émanation des intérêts de tous, et leur faiblesse, car ces intérêts doivent être filtrés par des principes de justice acceptables pour les électeurs à partir de leur conception du bien. (Audard, 2002, p. 62)

Rawls reconnaît ce défi posé par le pluralisme raisonnable. De plus, puisqu'il admet volontiers la possibilité que les utilitaristes fassent partie d'un consensus par recoupement sur le choix des principes de justice de la théorie de la justice comme équité, il n'exclut pas systématiquement les raisons utilitaristes. Ce qu'il exclut, c'est que seules des raisons utilitaristes puissent prévaloir dans le débat public. De plus, les raisons utilitaristes internes doivent, pour être valides, devenir des raisons publiques.

Par ailleurs, j'ai déjà souligné que l'utilitarisme se présente aujourd'hui sous de multiples aspects. La version mise de l'avant par John Harsanyi que j'ai résumée dans le précédent chapitre me semblait être particulièrement sophistiquée en tant que théorie morale

et pertinente pour évaluer le choix de principes de justice. Parlant de l'utilitarisme d'Harsanyi, qui rappelons-le ne s'intéresse pas à toutes les préférences individuelles mais seulement aux préférences morales qu'aurait un spectateur impartial doué de sympathie, Audard poursuit :

L'utilitarisme ne se limite donc pas aux raisons internes, il s'efforce de voir comment elles deviennent valides en tant que raisons publiques. Il fournit un mode de raisonnement et d'argumentation pour le débat public et la décision. (Audard, 2002, p. 61)

Les raisons utilitaristes ont donc leur place dans le débat public, contrairement à ce que la critique de Rawls peut sembler soutenir si on ne lui apporte pas quelques nuances. Mais finalement, en considérant tout l'éventail des conceptions politiques que l'on retrouve de nos jours, on peut se demander si la théorie de Rawls et l'utilitarisme sont véritablement des adversaires ou des alliés. Quand on considère leurs implications pour les politiques sociales, on constate par exemple que les implications de la théorie de Rawls et celle d'Harsanyi ne sont finalement pas si éloignées l'une de l'autre. J'estime également que l'application du principe d'utilité restreinte commanderait des politiques somme toute fort similaires à celles qu'exigerait l'application du principe de différence, malgré les différences subtiles qui existent bel et bien entre les deux. Selon Audard, les vrais adversaires de Rawls et de l'utilitarisme « *sont ailleurs, soit dans une conception de la décision politique détachée de toute visée normative morale et vouée à la seule efficacité, soit encore dans un élitisme technocratique qui prétend connaître le « bien du peuple » sans avoir à le consulter.* » (Audard, 2002, p. 62). Ceci m'amène maintenant, dans le cadre de la prochaine section, à considérer quelques affinités et éléments de désaccords entre l'utilitarisme et la théorie de Rawls.

3.8 Affinités et points de désaccords entre la théorie de Rawls et l'utilitarisme

De toute évidence, il existe certaines affinités entre la théorie de Rawls et l'approche utilitariste. Dans un texte portant sur Rawls et l'utilitarisme (Scheffler, 2003, p. 426-459), Samuel Scheffler identifie trois grandes similarités générales entre les deux théories. Premièrement, la théorie de Rawls partage avec l'utilitarisme le souhait d'élaborer une

conception de la justice ayant un caractère systématique et constructiviste, deux qualités indéniables de la pensée utilitariste aux yeux de Rawls. En fait, l'incapacité des théories rivales à offrir comme solution de rechange à l'utilitarisme une analyse aussi systématique de la justice explique en bonne partie, selon Rawls, le succès et la persistance historique de la philosophie utilitariste. Dans la préface de *Théorie de la Justice*, Rawls souligne en effet l'importance d'opposer à l'utilitarisme « *une conception morale systématique et applicable* » (Rawls, 1987, p. 20). La deuxième similarité que voit Scheffler a trait au fait que la théorie de Rawls subordonne les préceptes du sens commun concernant la justice à des principes supérieurs d'une manière analogue à celle qui permet à l'utilitarisme de dériver tout précepte de ce type à partir du seul et unique critère de la maximisation de l'utilité. Ces deux premiers points d'accord entre l'utilitarisme et la théorie de Rawls leur permettent de se positionner toutes deux contre une approche intuitionniste de la philosophie morale. La troisième similarité identifiée par Scheffler, l'affirmation du caractère holistique de la justice distributive, place l'utilitarisme aussi bien que la théorie de Rawls en opposition avec le libertarisme de Robert Nozick. Dans *Anarchie, État et utopie*, Nozick critique d'ailleurs les deux théories sur ce point lorsqu'il argue que le principe utilitariste aussi bien que les principes de justice de Rawls sont tous deux des principes de distribution organisés en modèle et que, par conséquent, ils ne respectent pas la liberté des individus. (Nozick, 1988, p. 195-205) Par cela, il entend que ces principes exigent que le résultat d'une distribution juste respecte une certaine configuration prédéterminée. À ce type de principe de justice distributive organisé en modèle, Nozick oppose son principe historique d'habilitation naturelle. Ainsi, alors que les théories de Rawls et de Harsanyi concentrent leur attention à une analyse macro de la justice, la théorie de Nozick opte pour une perspective micro.

Par ailleurs, Scheffler identifie également deux points de désaccords majeurs entre la théorie de Rawls et la doctrine utilitariste. Tout d'abord, la philosophie utilitariste, avec sa méthode agrégative, ne traite pas le caractère individuel et séparé des personnes de la même façon que la théorie de Rawls, qui accorde une plus grande importance au concept de réciprocité. Il s'ensuit, et c'est là le principal reproche de Rawls à l'égard de l'utilitarisme, que la maximisation de l'utilité peut imposer indûment des sacrifices inacceptables à certains individus au bénéfice de certains autres. Dans la position originelle, ce rejet de toute logique

agrégative et sacrificielle se traduit par l'interdiction pour les partenaires d'avoir recours à toute forme de raisonnement probabiliste. Comme je l'ai expliqué dans le chapitre précédent, c'est donc là que l'on retrouve la nature du désaccord entre Rawls et Harsanyi quant à ce qui constitue le choix rationnel à faire derrière le voile d'ignorance.

Le second point de désaccord principal entre Rawls et l'utilitarisme, qui n'est pas étranger au premier, est lié à la conception pluraliste du bien que défend Rawls, en opposition avec la conception moniste implicite à toute théorie utilitariste. L'utilitarisme subordonne le juste au bien, ce qui en fait une doctrine téléologique aux yeux de Rawls. L'utilitarisme définit effectivement la justice en fonction d'un but prédéterminé, celui de la maximisation de l'utilité. Ainsi, dans la pensée utilitariste, la justice est au service du bien-être de l'ensemble des individus qui composent la société. En ce sens, c'est une doctrine résolument conséquentialiste basée, d'un point de vue méthodologique, sur un profond individualisme. Cela étant dit, il ne faut pas confondre ici individualisme méthodologique et égoïsme⁴. L'utilitarisme n'affirme nullement que les individus ne se soucient pas du sort des autres. Il postule simplement que dans la détermination de ce qui est juste socialement, on doit procéder à un calcul dans lequel on fait la somme des utilités individuelles des membres de la société. C'est donc le sort des individus pris isolément qui constitue la base fondamentale de l'analyse utilitariste et « *l'intérêt collectif n'est rien d'autre que la somme des intérêts individuels.* » (Arnsperger et Van Parijs, 2003, p. 17)

D'autre part, en ce qui concerne le caractère téléologique de l'utilitarisme, Brian Barry souligne à juste titre que le critère utilitariste, s'il représente bel et bien un objectif ou un but défini préalablement, n'identifie pas pour autant une cible fixe à atteindre mais plutôt une flèche nous indiquant la voie à suivre. (Barry, 1989, p. 308) En effet, tout ce que le principe d'utilité nous dit, c'est que plus d'utilité est mieux que moins; il n'y a pas de somme déterminée d'utilité qu'une société doit atteindre pour être considérée juste. De plus, rappelons-le, les versions modernes de l'utilitarisme ont pour la plupart abandonné la

⁴ Dans un certain sens, l'utilitarisme peut même être considéré comme étant anti-individualiste puisque, en bout de ligne, l'«*intérêt collectif l'emporte toujours sur l'intérêt particulier de chacun.*» (Arnsperger et Van Parijs, 2003, p. 17)

définition expérientialiste de l'utilité pour adopter une définition préférentialiste; il ne s'agit donc plus de maximiser la différence entre les plaisirs et les peines mais plutôt de satisfaire tant que faire se peut les préférences des individus. Quand, en plus, on stipule comme Harsanyi qu'une théorie morale utilitariste se restreint aux préférences morales (impartiales), il demeure alors vrai que le juste est défini en fonction du bien mais ce qui est entendu par « le bien » est défini impartialement de manière très subtile et nuancée.

3.9 Différentes conceptions de la rationalité

Avant de conclure ce chapitre, j'aimerais ajouter un dernier point concernant les différentes conceptions de la rationalité sur lesquelles peut se baser une théorie éthique. Au sein de la doctrine utilitariste, la rationalité à laquelle on fait référence est purement instrumentale; elle commande à l'individu de mettre en œuvre les meilleurs moyens pour arriver à ses fins. Le recours à une conception instrumentale de la rationalité dans l'élaboration d'une théorie morale afin d'expliquer ce qui motive un individu à agir moralement implique que les règles morales peuvent toujours être interprétées en termes d'impératifs hypothétiques, comme le suggère Harsanyi dans un article datant de 1958 intitulé *Ethics in Terms of Hypothetical Imperatives*. Harsanyi explique en outre ceci :

Si les règles morales sont toutes des impératifs hypothétiques, ce sont donc des conseils et non des exigences. [...] Ce sont des conseils pour des personnes qui ont déjà une attitude morale, leur disant quelle sorte de comportement est compatible avec l'attitude morale qu'ils entretiennent. (Harsanyi, 1958, p. 308)

Puisqu'il présuppose que les partenaires sont dotés d'un sens de la justice et parce que de nombreux passages de la première partie de *Théorie de la Justice* présentent la rationalité des partenaires de la position originelle comme étant une rationalité instrumentale, on peut être tenté d'interpréter les principes de justice de Rawls de cette façon. En fait, une justification utilitariste indirecte des principes de justice rawlsiens passe nécessairement par une telle interprétation. Or, pour Rawls, le raisonnable encadre le rationnel de manière stricte et les individus moraux ont l'obligation d'agir selon les principes de justice choisis, qu'ils le souhaitent ou non. La théorie déontologique de Rawls ne saurait donc traiter les règles

morales comme des impératifs hypothétiques; ce sont nécessairement des impératifs catégoriques pour des êtres rationnellement autonomes.

Par ailleurs, les modèles utilitaristes s'appuient généralement sur une pure rationalité idéale. Autrement dit, ils font l'hypothèse d'une pleine et parfaite rationalité imputée aux agents. Pour être logiquement cohérentes, les préférences de l'agent pleinement rationnel doivent se conformer à différents axiomes, dont j'ai mentionné les plus importants précédemment : complétude et continuité de la relation de préférences, équivalence probabiliste, indépendance des alternatives pertinentes. La principale critique faite à la conception idéale de la rationalité est que le comportement observé des êtres humains ne s'y conforme généralement pas. Or, qualifier le comportement humain d'irrationnel ne nous avancerait à rien. Ainsi, on préfère dire que le comportement de l'être humain, sans être pleinement rationnel, est tout de même modérément rationnel. À la rationalité maximale proposée par la conception idéale, Christopher Cherniak oppose le concept de rationalité minimale. Selon cette conception minimale de la rationalité, un individu rationnel entreprend certaines des actions (mais pas nécessairement toutes les actions) qui sont apparemment appropriées en fonction de ses désirs et de ses croyances. (Cherniak, 1986, p. 9) Herbert Simon avance quant à lui l'idée d'une rationalité limitée qui, contrairement à la conception idéale, tient compte des limites associées aux capacités mentales de l'être humain. Selon Simon, les limites psychologiques de l'être humain en ce qui a trait notamment à ses capacités de calcul et ses habiletés prédictives font en sorte que la rationalité humaine n'est qu'une approximation grossière et simplifiée du genre de rationalité impliquée dans la plupart des modélisations du choix rationnel. (Simon, 1955, p. 101) Une des conséquences du caractère limité de la rationalité humaine est que l'agent se contente généralement d'un certain niveau minimal de satisfaction et d'aspiration plutôt que de tenter de maximiser son utilité. En quelque sorte, l'optimisation est limitée par les coûts cognitifs de la délibération car, n'étant pas omniscients (comme le serait un agent idéalement rationnel), nous devons acquérir toutes les informations pertinentes et procéder à des calculs complexes pour faire un choix éclairé, ce qui demande temps et effort. Or, ces ressources ne sont pas illimitées et, tout compte fait, il ne serait paradoxalement pas avantageux d'adopter une attitude orientée directement vers la pleine maximisation.

La rationalité humaine limitée de Simon est un concept différent du concept classique de rationalité substantielle puisqu'il s'agit d'une rationalité procédurale qui ne détermine pas de contenu sémantique fixe au principe de rationalité. Ainsi, à défaut d'être pleinement rationnels et de se conformer aux axiomes de la rationalité idéale, les êtres humains raisonnent selon des schèmes de pensée qui peuvent tout de même être qualifiés de rationnels. Selon cette conception procédurale, ce ne sont pas tant les choix que font les individus qui servent à déterminer le caractère rationnel de leurs actions mais plutôt la manière, le processus par lequel ces choix sont faits. Il semble clair que la rationalité dans la conception défendue par Rawls n'est ni substantielle ni idéale mais plutôt procédurale et limitée car, comme nous avons pu le voir, le choix des partenaires dans la position originelle ne respecte pas tous les axiomes bayésiens de la rationalité. La manière dont ce choix est fait et la sélection des raisons qui le justifie constituent en effet le cœur de l'argumentation de Rawls. On pourrait en dire bien davantage sur les différentes conceptions de la rationalité et leurs implications dans l'élaboration d'une théorie de la justice, mais, encore une fois, le temps et l'espace me manquent pour mener à bien cette tâche dans le cadre du présent mémoire.

3.10 Conclusion

La relation qu'entretient la théorie de Rawls avec l'utilitarisme n'est pas simple. Ainsi, alors que l'un des objectifs avoués de *Théorie de la Justice* était de présenter une puissante critique de la doctrine utilitariste, nous avons pu voir que les écrits ultérieurs de Rawls viennent nuancer sa position vis-à-vis l'utilitarisme. Par ailleurs, l'évolution de la pensée de Rawls et son insistance sur le concept de consensus par recoupement ne peuvent expliquer seuls toutes les difficultés et subtilités inhérentes à une comparaison entre la théorie de Rawls et l'utilitarisme. En effet, le développement de nouvelles versions sophistiquées de la doctrine utilitariste viennent également compliquer la tâche. Cela m'a d'ailleurs amené à faire quelques précisions et distinctions importantes pour clarifier certaines questions. La distinction entre des principes utilitaristes et l'utilitarisme philosophique en constitue un bon exemple. Celle entre la procédure de décision utilisée et le résultat de la procédure en est une

autre, tout comme celle entre un critère du choix rationnel et un principe de justice. La précision concernant le caractère semi-contractualiste de la théorie de Rawls est également très pertinente, tout comme la discussion sur les deux niveaux de justification. Finalement, la différence entre motivations individuelles et raisons morales est elle aussi très utile.

Ce qu'il faut retenir, c'est que, malgré certaines apparences, la théorie de Rawls n'est pas utilitariste mais bel et bien contractualiste. Rawls refuse d'appuyer sa conception de la justice sur la thèse fondamentale de l'utilitarisme philosophique, qui veut que seuls les faits liés au bien-être individuel aient une pertinence morale. Cela dit, dans ses textes publiés après *Théorie de la justice*, Rawls insiste pour dire que son objectif n'est pas de réfuter l'utilitarisme philosophique mais plutôt de soutenir que le contractualisme fournit un meilleur cadre pour élaborer une conception politique de la justice dans le contexte du pluralisme raisonnable. La possibilité de rallier les utilitaristes autour des deux principes de justice devient alors particulièrement intéressante, d'autant plus que l'utilitarisme fournit un certain nombre de raisons pertinentes pour appuyer les principes rawlsiens. Dans la plupart des pays démocratiques, utilitaristes et rawlsiens apparaissent ainsi généralement davantage comme des alliés que des adversaires lorsque vient le temps de s'entendre sur les politiques sociales à adopter.

CONCLUSION

Nous avons vu que la théorie de la justice comme équité de John Rawls énonce deux principes de justice s'appliquant au fonctionnement des institutions sociales de base d'une société démocratique contemporaine. Le premier stipule que les individus devraient se voir garantir un système cohérent de droits et libertés le plus étendu possible. Le second affirme que, dans le cadre d'une juste égalité des chances, les inégalités économiques devraient être tolérées uniquement dans la mesure où elles permettent d'améliorer le sort du groupe le plus défavorisé de la société. La théorie de Rawls se veut une conception politique de la justice applicable à une société pluraliste et démocratique. Dans ce contexte, les principes de justice rawlsiens doivent faire l'objet d'un consensus par recoupement ou, autrement dit, être approuvés et soutenus par un ensemble d'individus ayant des conceptions du bien différentes mais raisonnables. La justification des principes de justice dans la théorie de Rawls est ainsi une justification publique qui se fonde dans le concept du contrat social. Pour découvrir le contenu de ces principes, Rawls propose de les dériver d'une construction artificielle, celle de la position originelle, au sein de laquelle un individu représentatif ignorant la position sociale qu'il occupe choisit les principes les plus à même d'avancer les intérêts des personnes qu'il représente. La décision que prend l'individu représentatif est alors comprise comme un choix rationnel fait en situation d'incertitude. Cependant, les informations mises à sa disposition ainsi que les raisons justifiant son choix sont limitées par des considérations contractualistes qui modèlent les circonstances de la position originelle. La théorie du choix rationnel occupe donc une place importante dans la théorie de Rawls, mais il serait faux de dire que ses principes de justice en sont dérivés directement. Il est en effet plus exact de dire que la théorie de la justice comme équité fait usage de la théorie du choix rationnel d'une manière bien précise et dans le contexte très particulier de la position originelle. Par ailleurs, il ne faut jamais perdre de vue que la position originelle est construite de façon à représenter une situation initiale d'égalité entre individus libres et rationnels souhaitant s'entendre sur les termes d'une coopération sociale mutuellement avantageuse. Ainsi, les principes justes pour

régir le fonctionnement des institutions sociales sont ceux sur lesquels s'entendraient des personnes libres et rationnelles si elles étaient placées dans le contexte hypothétique d'égalité que constitue la position originelle.

John Harsanyi s'accorde avec Rawls pour dire que les règles morales encadrant la vie sociale devraient être choisies de manière impartiale par un individu ne sachant pas quelle position il occupe dans la société. Toutefois, Harsanyi estime que le choix rationnel d'un individu placé dans une telle situation est d'opter pour le principe d'utilité moyenne et non pour les deux principes de justice rawlsiens. Son raisonnement est simple : ne sachant pas quelle position il occupe, l'individu rationnel suppose qu'il a une chance égale de se retrouver dans chacune de positions possibles et, puisqu'il tente de maximiser son utilité espérée, il choisit le principe qui commande de maximiser l'utilité moyenne des personnes qu'il pourrait devenir. L'argumentation d'Harsanyi repose alors sur une application fructueuse de la théorie bayésienne de la décision rationnelle, qui suppose une interprétation logique du concept de probabilité. Rawls, quant à lui, interdit toute forme de raisonnement probabiliste dans la position originelle. Les deux théories sont ainsi en opposition sur ce point bien précis. D'un point de vue technique, c'est ce qui explique pourquoi l'une aboutit à un principe utilitariste et pas l'autre. Cela dit, si le principe d'utilité moyenne est évidemment un principe utilitariste, il n'est pas clair que la justification du choix de ce principe proposée par Harsanyi soit elle-même strictement utilitariste. En effet, il flirte avec le contractualisme lorsqu'il écarte du calcul utilitariste certaines préférences jugées antisociales. Il ne donne pas de raison utilitariste pour expliquer cette prise de position, ce qui crée une distance évidente entre sa théorie et la thèse fondamentale de l'utilitarisme philosophique.

En outre, si l'utilitarisme peut être décrit de manière très générale comme *la morale de l'utilité sociale*, l'analyse d'Harsanyi correspond-elle davantage à ce qu'il convient d'appeler une étude de *l'utilité sociale de la morale* ? Les deux expressions ne sont pas équivalentes, loin de là; tandis qu'une morale de l'utilité est résolument normative, une étude de l'utilité de la morale est d'abord et avant tout descriptive. Plus encore, si la morale étudiée n'est pas une morale conséquentialiste fondée sur le principe d'utilité mais plutôt une morale déontologique fondée sur des bases contractualistes, il devient difficile de soutenir que l'on

est en présence d'une théorie morale véritablement utilitariste. La question qui se pose est donc la suivante : selon quels critères une théorie morale peut-elle être qualifiée d'utilitariste? Ou encore : à partir de quel moment une théorie morale qui a recours au concept d'utilité devient-elle utilitariste ou cesse-t-elle de l'être? Dans tous les cas, une définition précise du terme *utilité* ainsi qu'une énumération des caractéristiques essentielles de toute théorie morale utilitariste digne de ce nom s'avèrent indispensables pour répondre adéquatement à ces questions. Je n'ai pas la prétention d'avoir accompli ces tâches en entier dans le cadre de ce mémoire, mais j'ai tout de même l'espoir d'avoir contribué à clarifier certaines idées essentielles reliées à ce thème.

Par ailleurs, même si certains éléments de l'argumentation de Rawls à partir de la position originelle ont des allures utilitaristes, la justification de la conception politique de la justice comme équité qu'il met de l'avant n'est pas utilitariste. Contrairement à ce qu'affirme Sydney Alexander, la théorie de Rawls n'est donc pas un utilitarisme indirect; il s'agit bel et bien d'une théorie déontologique contractualiste. Le point de vue moral privilégié par Rawls n'est pas celui d'un spectateur impartial dont l'objectif est de maximiser la satisfaction des préférences individuelles. Selon Rawls, le rôle des institutions sociales n'est pas d'administrer efficacement les ressources rares au plus grand avantage de la somme des intérêts individuels mais plutôt d'assurer l'encadrement d'un système de coopération fondé sur les idées d'avantage mutuel et de réciprocité. L'idée d'adhésion volontaire de tous les citoyens à la conception de la justice mise de l'avant est donc primordiale pour la théorie de Rawls, ce qui n'est pas le cas pour l'utilitarisme.

Cependant, la manière dont Rawls ramène le contenu d'un accord entre plusieurs personnes aux intérêts divergents au choix rationnel d'un seul individu représentatif peut créer une certaine confusion sur ce point. Bien entendu, les contraintes imposées aux informations dont dispose l'individu représentatif et la limitation des raisons sur lesquelles peut s'appuyer son choix garantissent le caractère équitable des principes retenus, reflétant ainsi l'idée de réciprocité implicite dans le concept d'accord volontaire. J'insiste tout de même pour souligner que cette façon indirecte de procéder brouille inutilement les arguments de Rawls contre l'utilitarisme. Comme le suggère Scanlon, je considère qu'une

argumentation plus directe en faveur des deux principes de justice rawlsiens servirait mieux la justification contractualiste que recherche Rawls. En effet, Rawls aurait pu simplement dire que les principes justes sont les principes que personne ne pourrait raisonnablement rejeter, soumettant ainsi le contenu de l'accord à une négociation rationnelle soumise à des contraintes raisonnables empêchant quiconque de profiter d'un pouvoir de négociation inégal. L'option privilégiée par Rawls crée par ailleurs l'occasion d'analyser l'utilisation de la théorie du choix rationnel dans le cadre d'une théorie de la justice, ce qui n'est certes pas sans intérêt malgré les difficultés que cela comporte.

Il est en outre important de préciser que si la théorie morale de Scanlon généralise le contractualisme, la théorie de la justice de Rawls n'est uniquement qu'une conception politique de la justice valable pour une société démocratique libérale caractérisée par le contexte bien précis du pluralisme raisonnable. Puisque cette conception de la justice doit faire l'objet d'un consensus par recoupement impliquant toutes les doctrines morales raisonnables, les utilitaristes doivent pouvoir y adhérer également. Une justification utilitariste du choix des principes de justice rawlsiens doit donc être envisageable. Cette justification prend la forme d'un utilitarisme indirect dont l'objet est un ensemble de règles constitutives d'une association entre personnes libres et rationnelles. Pour cela, il faut dans un premier temps supposer que les préférences réelles de tous les individus, ou du moins de la très grande majorité, sont telles qu'ils souhaitent tous se voir garantir un système de droits et libertés le plus étendu possible. Pour justifier le choix du principe de différence, il faut cependant faire des hypothèses plus particulières sur la psychologie morale des individus. Mon opinion est qu'il est fort plausible que les préférences de la plupart des gens sont telles qu'ils opteraient plutôt pour un minimum social garantissant des conditions de vie décentes auquel serait juxtaposé le principe d'utilité moyenne.

Rawls lui-même admet volontiers que l'on puisse modéliser les préférences de l'individu représentatif dans la position originelle à l'aide d'une fonction d'utilité. Toutefois, cette fonction d'utilité ne représente pas les préférences d'un individu moyen ou médian, mais incorpore plutôt toutes les considérations qui servent à élaborer la théorie de la justice comme équité. Ces considérations reposent sur des bases contractualistes explicites mais,

rappelons-le, elles se veulent neutres à l'égard de toutes les doctrines morales raisonnables. Le concept d'utilité ici utilisé doit en outre être compris strictement comme une mesure de la satisfaction des préférences. Le critère de la maximisation de l'utilité attendue n'est alors pas un principe de justice en tant que tel mais plutôt un critère du choix rationnel en situation d'incertitude; il implique que le ou les principes choisis doivent maximiser l'utilité attendue, mais pas nécessairement que le principe voulant que l'on tente continuellement de maximiser directement l'utilité moyenne soit choisi.

De manière générale, je crois que toute tentative de dériver une théorie de la justice exclusivement à partir de la théorie du choix rationnel est vouée à l'échec. Le problème découle du fait que le concept de rationalité représente un puissant outil analytique pour expliquer le raisonnement humain mais ne peut à lui seul justifier moralement une décision. Ainsi, on peut utiliser la théorie du choix rationnel pour mettre en lumière certains aspects d'une théorie de la justice, mais on ne pourrait pas fonder la moralité uniquement à l'aide de celle-ci. En soi, une utilisation judicieuse de la théorie du choix rationnel est neutre d'un point de vue moral; elle nous dit simplement que des principes seront choisis par un individu rationnel si et seulement si ce sont ceux qui ont la plus grande espérance d'utilité à ses yeux mais elle ne détermine par elle-même aucun contenu spécifique à des principes dits justes. Si l'on n'insuffle pas une certaine dose de moralité au départ d'une analyse de la morale en terme de choix rationnel, on ne peut espérer que cette analyse aboutira sur des principes pouvant être qualifiés de moraux. C'est ce qui explique qu'Harsanyi est forcé de limiter les préférences devant être considérées dans le calcul utilitariste. Malheureusement, il ne justifie pas explicitement l'imposition de ces limites. Les bases philosophiques qui fondent la théorie morale d'Harsanyi demeurent ainsi mystérieuses.

Le projet de Rawls peut quant à lui être vu comme une tentative de donner de la substance à des principes qui seraient autrement trop vagues et difficilement applicables. Car en effet, dire que des principes justes sont ceux que préférerait un individu rationnel placé dans certaines circonstances bien précises ne suffit pas; il faut également expliciter les raisons qui justifient moralement ce choix. Bien que Rawls utilise la théorie du choix rationnel dans l'élaboration de sa théorie de la justice, cette dernière se fonde sur une justification

contractualiste adaptée au contexte particulier du pluralisme raisonnable. C'est ce cadre contractualiste et les raisons qu'il admet qui permet de déterminer le contenu spécifique des deux principes et non la théorie du choix rationnel, qui en serait incapable à elle seule. Cela dit, il existe entre la théorie de la décision rationnelle individuelle et l'utilitarisme une certaine affinité naturelle, ce qui rend pour le moins difficile et complexe l'argumentation indirecte de Rawls à partir de la position originelle. Je ne prétends pas que cette argumentation est invalide, mais plutôt qu'il ne s'agit peut-être pas de la meilleure stratégie argumentative pour défendre sa conception de la justice comme équité et surpasser l'utilitarisme.

Par ailleurs, j'insiste sur le fait que, dans le contexte qui prévaut actuellement dans la plupart des sociétés démocratiques, utilitaristes et défenseurs de la théorie de la justice comme équité peuvent former une alliance pour réclamer la mise en place de politiques sociales progressistes. Cette possibilité découle du fait que malgré certaines nuances philosophiques qui permettent de distinguer les deux théories, leurs implications pratiques ne sont pas très différentes. À tout le moins, elles sont plus semblables que celles des conceptions libertariennes ou communautaristes, par exemple. Par ailleurs, les différences qui demeurent permettent de rendre compte du défi que pose le fait du pluralisme raisonnable, c'est-à-dire d'essayer de trouver un point de rencontre entre les motivations individuelles et les raisons morales d'agir.

RÉFÉRENCES

- Alexander, S. 1974. « Social Evaluation through Notional Choice », *The Quarterly Journal of Economics*, Vol 88, p. 597-624.
- Arnsperger, C. et P. Van Parijs. 2003. *Éthique économique et sociale*, Éditions La Découverte.
- Barry, B. 1989. *Theories of Justice*, University of California Press.
- Brandt, R. 1936. « Utilitarianism Revised », *Mind*, Vol. 45, p. 137-156.
- Bernouilly, D. 1738. « Exposition of a New Theory on the Measurement of Risk », traduit et réimprimé dans *Econometrica*, Vol. 22 (1954), p. 23-26.
- Cherniak, C. 1986. *Minimal Rationality*, MIT Press.
- Fleming, M. 1952. « A Cardinal Concept of Welfare », *Quarterly Journal of Economics*, Vol. 66, p. 366-384.
- Gauthier, D. 2000. *Morale et contrat*, Mardaga.
- Harsanyi, J. 1953. « Cardinal Utility in Welfare Economics and the Theory of Risk-Taking », *Journal of Political Economy*, Vol. 61, p. 434-435.
- . 1955. « Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility », *Journal of Political Economy*, Vol. 63, p. 309-321.
- . 1958. « Ethics in Terms of Hypothetical Imperatives », *Mind*, Vol. 67, p. 305-316.
- . 1961. « On the Rationality Postulates Underlying the Theory of Cooperative Games », *Journal of Conflict Resolution*, Vol. 5, p. 179-196.
- . 1975. « Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory », *The American Political Science Review*, Vol. 69, p. 594-606.
- . 1982. « Morality and the theory of Rational Behavior ». In *Utilitarianism and Beyond*, sous la dir. de A. Sen et B. Williams, p. 39-62. Cambridge University Press.

- . 1985. « Does Reason Tell Us What Moral Code to follow and, Indeed, to follow Any Moral Code at All? », *Ethics*, Vol. 96, p. 42-55.
- Knowles, D. 2001. *Political Philosophy*, McGill-Queen's University Press.
- Lyons, D. 1972. « Rawls versus Utilitarianism », *The Journal of Philosophy*, Vol. 69, p. 535-545.
- Nootens, G. 1996. *La nature de la complémentarité entre le raisonnable et le rationnel chez Rawls*, Cahiers de l'ERES, No. 9603, Université du Québec à Montréal.
- Nozick, Robert. 1988. *Anarchie, état et utopie*, Presses Universitaires de France.
- Pogge, T. 1989. *Realizing Rawls*, Cornell University Press.
- RAWLS, J. 1955. « Two Concepts of Rules », *Philosophical Review*, Vol. 64, p. 3-32.
- . 1958. « Justice as Fairness », *Philosophical Review*, Vol. 67, p. 164-194.
- . 1963. « The Sense of Justice », *Philosophical Review*, Vol. 72, p. 281-305.
- . 1987. *Théorie de la justice*, Éditions du Seuil.
- . 1993. *Justice et démocratie*, Éditions du Seuil.
- . 1995. *Libéralisme politique*, Presse Universitaires de France.
- . 2004. *La justice comme équité : Une reformulation de Théorie de la justice*, Boréal.
- Scanlon, T. 1982. « Contractualism and Utilitarianism ». In *Utilitarianism and Beyond*, sous la dir. de A. Sen et B. Williams, p. 103-127. Cambridge University Press.
- Scheffler, S. 2003. « Rawls and Utilitarianism ». In *The Cambridge Companion to Rawls*, sous la dir. de S. Freeman, p. 426-459. Cambridge University Press.
- Sibley, W. M. 1953. « The Rational Versus the Reasonable », *The Philosophical Review*, Vol. 62, p. 554-560.
- Simon, H. 1955. « A Behavioral Model of Rational Choice », *Quarterly Journal of Economics*, Vol. 89, p. 99-118.
- Varian, H. 1997. *Introduction à la microéconomie*, De Boeck.
- Vergara, F. 2005. *Utilitarisme*, en ligne : <http://www.fvergara.com/utilitarisme.pdf>

Von Neumann, J. et O. Morgenstern. 1944. *Theory of Games and Economic Behavior*, Wiley.

BIBLIOGRAPHIE SÉLECTIVE

Livres et recueils de John Rawls

Rawls, John. *Théorie de la justice*, Éditions du Seuil. 1987. 665 p.

———. *Justice et démocratie*, Éditions du Seuil. 1993. 385 p.

———. *Libéralisme politique*, Presse Universitaires de France. 1995. 450 p.

———. *La justice comme équité : Une reformulation de Théorie de la justice*, Boréal. 2004. 283 p.

Articles de John Rawls

Rawls, John. « Outline of a Decision Procedure for Ethics », *Philosophical Review*, Vol. 60 (1951), p.177-197.

———. « Two Concepts of Rules », *Philosophical Review*, Vol. 64 (1955), p. 3-32.

———. « Justice as Fairness », *Philosophical Review*, Vol. 67 (1958), p. 164-194.

———. « The Sense of Justice », *Philosophical Review*, Vol. 72 (1963), p. 281-305.

———. « Some Reasons for the Maximin Criterion », *The American Economic Review*, Vol. 64 (1974), p. 141-146.

———. « Reply to Alexander and Musgrave », *Quarterly Journal of Economics*, Vol. 88 (1974), p. 633-655.

Livres et recueils sur John Rawls et son oeuvre

Freeman, Samuel (dir. publ.). *The Cambridge Companion to Rawls*, Cambridge University Press. 2003. 585 p.

———. *Rawls*, Routledge. 2007. 550 p.

———. (dir. publ.). *Lectures on the History of Political Philosophy / John Rawls*, Belknap Press. 2007. 476 p.

Kukhatas, Chandran & P. Petit. *A Theory of Justice and its Critics*, Stanford University Press. 1990. 169 p.

Pogge, Thomas. *Realizing Rawls*, Cornell University Press. 1989. 296 p.

———. *John Rawls: His Life and Theory of Justice*, Oxford University Press. 2007. 240 p.

Articles sur John Rawls et son œuvre

Alexander, Sydney. « Social Evaluation through Notional Choice », *The Quarterly Journal of Economics*, Vol 88 (1974), p. 597-624.

Brock, Dan. « The Theory of Justice », *University of Chicago Law Review*, Vol. 40 (1973), p. 486-499.

Chapman, John. « Rawls's Theory of Justice », *The American Political Science Review*, Vol. 69 (1975), p. 588-593.

Dworkin, Ronald. « The Original Position », *University of Chicago Law Review*, Vol. 40 (1973), p. 500-533.

Hare, R. M. « Rawls's Theory of Justice », *The Philosophical Quarterly*, Vol. 23 (1973), p. 144-155.

Lyons, David. « Rawls versus Utilitarianism », *The Journal of Philosophy*, Vol. 69 (1972), p. 535-545.

Musgrave, Richard A. « Maximin, Uncertainty and the Leisure Trade-off », *The Quarterly Journal of Economics*, Vol. 88 (1974), p. 625-632.

Nagel, Thomas. « Rawls on Justice », *The Philosophical Review*, Vol. 82 (1973), p. 220-234.

Nootens, Geneviève. *La nature de la complémentarité entre le raisonnable et le rationnel chez Rawls*, Cahiers de l'ERES, No. 9603 (1996), Université du Québec à Montréal. 27 p.

Scanlon, Thomas M. « Rawls's Theory of Justice », *University of Pennsylvania Law Review*, Vol. 121 (1973), p. 1020-1069.

Livre de John Harsanyi

Harsanyi, J. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge University Press. 1977. 352 p.

Articles de John Harsanyi

Harsanyi, John. « Cardinal Utility in Welfare Economics and the Theory of Risk-Taking », *Journal of Political Economy*, Vol. 61 (1953) p. 434-435.

———. « Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility », *Journal of Political Economy*, Vol. 63 (1955) p. 309-321.

———. « Approaches to the Bargaining Problem Before and After the Theory of Games », *Econometrica*, Vol. 24 (1956), p. 144-157.

———. « Ethics in Terms of Hypothetical Imperatives », *Mind*, Vol. 67 (1958), p. 305-316.

———. « On the Rationality Postulates Underlying the Theory of Cooperative Games », *Journal of Conflict Resolution*, Vol. 5 (1961), p. 179-196.

———. « A General Theory of Rational Behavior in Game Situations », *Econometrica*, Vol. 34 (1966), p. 613-634.

———. « Games with Incomplete Information Played by Bayesian Players », *Management Science*, Vol. 14 (1967-8), p. 159-182, p. 320-334, p. 486-502.

———. « Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory », *The American Political Science Review*, Vol. 69 (1975), p. 594-606.

———. « Does Reason Tell Us What Moral Code to follow and, Indeed, to follow Any Moral Code at All? », *Ethics*, Vol. 96 (1985), p. 42-55.

———. « Rule Utilitarianism and Decision Theory ». *Erkenntnis*, Vol. 11 (1977), p. 25-53.

———. « Basic Moral Decisions and Alternative Concepts of Rationality », *Social Theory and Practice*, Vol. 9 (1983), p. 231-243.

Article sur John Harsanyi et son œuvre

Brock, Horace. « A Critical Discussion on the Work of John C. Harsanyi », *Theory and Decision*, Vol. 9 (1978), pp. 349-367.

Autres documents pertinents

Arnsperger, Christian et Philippe Van Parijs. *Éthique économique et sociale*, Éditions La Découverte. 2003. 119 p.

Baier, Kurt. *The Moral Point of View; A Rational Basis for Ethics*, Cornell University Press. 1969. 326 p.

Barry, Brian. *Theories of Justice*, University of California Press. 1989. 428 p.

———. *Justice as Impartiality*, Clarendon Press. 1995. 315 p.

Brandt, Richard. « Utilitarianism Revised », *Mind*, Vol. 45 (1936), p. 137-156.

Bernouilly, Daniel. « Exposition of a New Theory on the Measurement of Risk », traduit et réimprimé en 1954 dans *Econometrica*, Vol. 22 (1738), p. 23-26.

Cherniak, Christopher. *Minimal Rationality*, MIT Press. 1986. 161 p.

Copp, David et David Zimmerman (dir. publ.). *Morality, Reason and Truth : New Essays on the Foundations of Ethics*, Rowman & Allanheld. 1984. 331 p.

Couture, Jocelyne (dir. publ.). *Éthique et rationalité*, Mardaga. 1992. 125 p.

Daniels, Norman. *Justice and Justification*, Cambridge University Press. 1996. 365 p.

Fleming, Marcus. « A Cardinal Concept of Welfare », *Quarterly Journal of Economics*, Vol. 66 (1952), p. 366-384.

Fleurbaey, Marc. *Théories économiques de la justice*, Economica. 1996. 250 p.

Gauthier, David. *Morale et contrat*, Mardaga. 2000. 430 p.

Knowles, Dudley. *Political Philosophy*, McGill-Queen's University Press. 2001. 392 p.

Mongin, Philippe. « L'optimisation est-elle un critère de rationalité individuelle? », *Dialogue*, Vol. 33 (1994), p. 191-222.

Moser, Paul K. (dir. publ.) *Rationality in Action*, Cambridge University Press. 1990. 493 p.

- Nozick, Robert. *Anarchie, état et utopie*, Presses Universitaires de France. 1988. 442 p.
- Parfit, Derek. *Reasons and Persons*, Oxford University Press. 1984. 543 p.
- Ramsey, Frank. *Logique, philosophie et probabilités*, Librairie philosophique J. Vrin. 2003. 352 p.
- Scanlon, Thomas M. *What We Owe to Each Other*, Belknap Press of Harvard University Press. 1998. 420 p.
- Sibley, W. M. « The Rational Versus the Reasonable », *The Philosophical Review*, Vol. 62 (1953), p. 554-560.
- Simon, Herbert. « A Behavioral Model of Rational Choice », *Quarterly Journal of Economics*, Vol. 89 (1955), p. 99-118.
- Sen, Amartya et Bernard Williams (dir. publ.). *Utilitarianism and Beyond*, Cambridge University Press. 1982. 290 p.
- Varian, Hal. *Introduction à la microéconomie*, De Boeck. 1997. 758 p.
- Vergara, Francisco. *Utilitarisme*. 2005. En ligne : <http://www.fvergara.com/utilitarisme.pdf>
- Von Neumann, John et Oskar Morgenstern. *Theory of Games and Economic Behavior*, Wiley. 1944. 641 p.